DOCUMENT RESUME

ED 185 118 TM 800 153

AUTHOR Holmes, Susan E.
TITLE ESEA Title I Linking Project. Final Report.
INSTITUTION Oregon State Dept. of Education, Salem.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Jan 80
CONTRACT 300-79-0019
NOTE 132p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Achievement Gains; Achievement Tests; Educational Testing; Elementary Education; Elementary School Mathematics; *Equated Scores; Federal Programs; Goodness of Fit; *Item Banks; Item Sampling; *National Norms; Norm Referenced Tests; Reading Achievement; Statistical Analysis; Test Norms
IDENTIFIERS California Achievement Tests; *Elementary Secondary Education Act Title I; Northwest Evaluation Association; *Rasch Model

ABSTRACT

The Rasch model for test score equating was compared with three other equating procedures as methods for implementing the norm referenced method (RMC Model A) of evaluating ESEA Title I projects. The Rasch model and its theoretical limitations were described. The three other equating methods used were: linear observed score equating, linear true score equating, and equipercentile equating. Data used in the study were obtained by testing 6,000 students in grades three through eight with the norm-referenced California Achievement Tests, Form C (CAT/C), reading and mathematics, and with tests made up from items from two Rasch calibrated item banks in reading and mathematics. The results of the four equating procedures were compared with actual observed score distributions on the CAT/C for each item bank raw score. All four equating techniques yielded similar results with the least accuracy at the low end of the score distributions. It appeared that use of the Rasch model was feasible, but the accuracy of the results it produced was questionable. (Author/CTM)

ESEA Title I Linking Project
Final Report

January 1980

Verne A. Duncan
State Superintendent of
Public Instruction

Oregon Department of Education
700 Pringle Parkway SE
Salem, Oregon 97310

ESEA Title I Linking Project
Final Report

The Title I Linking Project was a research effort jointly sponsored by the Oregon Department of Education and the United States Office of Education. The purpose of the project was to explore an alternative Title I evaluation reporting strategy; that is, the use of achievement tests developed from Rasch calibrated item banks and this report describes a methodology for such an approach. Tables are included which can be used to convert item bank achievement estimates into the Normal Curve Equivalent units necessary for Title I reporting. The utility of the tables is demonstrated by their application to real data.

Report By
Susan E. Holmes
Research Associate
Educational Program Audit Division

Acknowledgment

This report represents the efforts of a number of individuals. It would be impossible to acknowledge all those who have contributed, but I would like to mention a few to whom I feel especially indebted.

A large measure of support for this Project was provided by the Oregon Department of Education. Gordon Ascher and Marshall Herron provided assistance in formulating the Project and created and maintained an environment where research could be conducted.

The staff of two Oregon school districts, North Clackamas 12 and Oregon City 62, voluntarily supported the Project by providing data collection sites. Clackamas Education Service District staff provided test scoring services.

Louis Bashaw, Gary Marco, Robert Rentz, and Richard Woodcock helped to clarify many of the technical issues encountered throughout the Project. Portland School District 1J and its evaluation staff made available data and technical expertise in Rasch model technology.

Tom Haladyna was actively involved in the data analysis phase of the Project and contributed valuable suggestions, criticisms and unwaivering interest. Yuzn-Fure Suen provided statistical knowledge and computer know-how.

This report was prepared under Contract No. 300-79-0019 by the Department of Education of the State of Oregon for the United States Office of Education, Department of Health, Education and Welfare.

TABLE OF CONTENTS

INTRODUCTION

Since 1974, the United States Office of Education has attempted to implement Title I evaluation reporting procedures which are outcome oriented and provide for the aggregation of achievement data across all Title I projects. The system is intended to facilitate the collection and analysis of data in six areas: student participation, parent involvement, staffing, inservice training, cost and project impact. Major emphasis has been placed on the last of these, project impact, particularly for those projects providing instruction in the basic skill areas of reading, language arts and mathematics.

Three models or designs have been proposed for the evaluation of project impact, defined here in terms of achievement gains resulting from Title I intervention (Tallmadge and Wood, 1978). All three models may be used in conjunction with either a criterion-referenced or norm-referenced achievement test, and each involves the pretesting and posttesting of Title I students. Achievement gains are then determined and quantified in terms of a common metric called the Normal Curve Equivalent (NCE) scale.

Of particular interest here is Model A, referred to as the norm-referenced model. The basic assumption of this model is that the percentile rank of Title I students will remain constant in relation to the group on which the achievement test was normed from pretest to posttest if no Title I intervention occurred. Of the several proposed models, Model A is expected to be the most widely adopted since it utilizes procedures already in

place in many projects. In addition, Model A does not necessitate the identification of a local comparison group, since the no-treatment expectation is derived from normative data collected by the test publisher. As a result, in addition to being computationally less complex than the other models, Model A is less expensive to implement.

Model A may be implemented in several ways. The first, referred to as Model A1, involves administering a normed achievement test to Title I students at pretest and posttest time. A no-treatment expectation is then determined by assuming that these Title I students would have maintained their percentile rank with respect to the norm group from pretest to posttest if there had been no Title I intervention.

One of the problems encountered in the application of Model A1 is the availability of normed achievement tests which adequately reflect Title I curriculum at the local level. Since commercially developed instruments are generally designed to encompass a fairly broad range of goals and objectives, more specific local interests are frequently thinly represented. In addition, there is the danger that the selection of a normed achievement test which poorly reflects local objectives may adversely effect the course of future curriculum decisions.

Since the availability of appropriate normed achievement tests is a real concern, a second implementation plan for Model A, referred to as Model A2, has been proposed. This model involves the administration of a non-normed achievement test to Title I students at pretest and posttest time. In addition, a normed test is also administered at either of the two test times. By equating the normed and non-normed tests, normative data compiled for the normed test can be used to derive a no-treatment expectation for the non-normed test.

Model A2 does offer some recourse to those districts which are unable to locate a normed achievement test which adequately reflects their program objectives. However, should the project objectives change significantly enough to necessitate changes in the non-normed test, the entire equating process outlined in Model A2 would need to be repeated for the new non-normed test. The computational complexity of Model A2 coupled with the prospect of replicating the effort every few years is enough to make many local districts reluctant to consider this approach to Title I evaluation.

A third alternative for the implementation of Model A, here referred to as "Model A3," also exists. This approach attempts to provide local districts with more flexibility in selecting achievement tests for Title I evaluation, while eliminating the need for replicating the test equating effort each time a new test is selected or an existing one modified.

Model A3 involves the use of item banks developed through Rasch model item analysis techniques.

Over the past few years, interest in applying Rasch model technology in educational measurement has greatly increased. One of the advantages of this approach to measurement lies in the fact that, once a Rasch calibrated item bank has been constructed, all possible subsets of items drawn from the bank are automatically equated (i.e., result in test information on a common metric). In order to utilize a Rasch calibrated item bank in Title I evaluation reporting, normative information must also be available for these common metric scores. By determining the relationship between selected subsets of items drawn from the bank and a series of normed tests, the normative data compiled for the normed test can be associated with the scale underlying the item bank. Once this relationship is defined, any appropriately chosen subset of bank items could then be used to evaluate

project impact, provided the items were administered at the appropriate time of the year. Local districts could select one set of items to identify Title I eligible students, another to determine the no-treatment expectation (pretest) for this group, and yet another to evaluate growth (posttest). Since all possible subsets of items drawn from the bank are automatically equated, tests which more accurately reflect local curriculum can be constructed, with the test equating process being performed only once.

The purpose of the research reported here was to explore Model A3 as an alternative Model A implementation strategy for Title I evaluation reporting. This research effort, referred to as the Title I Linking Project, was jointly sponsored by the Oregon Department of Education and the United States Office of Education. The specific Project objectives were to:

1. Describe methodology for equating a Rasch calibrated item bank and a norm-referenced achievement test.
2. Evaluate the results obtained from a Rasch model approach to test equating with those of a more conventional approach.
3. Produce tables which permit the conversion of item bank test scores into NCE units.
4. Provide a prototype Title I evaluation report based on tests developed from a Rasch calibrated item bank.

Theoretical Considerations

The Rasch Model

Although latent trait models were introduced by Frederick Lord (1952, 1953) almost 25 years ago, interest in using them to solve educational measurement problems has been a relatively recent development. Latent trait theory supposes that test performance can be explained in terms of student characteristics which are called traits (Lord and Novick, 1968). Since these traits are not directly measurable, they are referred to as latent traits, or abilities, and it is typically assumed that only one trait underlies performance on any given test. Throughout this paper, the terms "latent trait", "ability" and "achievement" will be used interchangeably.

Several latent trait models, each defined in terms of the number of parameters required to describe the testing situation, have been studied. One of the more promising is the simple logistic, or Rasch, model. The primary advantage of this particular model over other latent trait models lies in its simplicity. It requires fewer parameters, thus simplifying analyses. In addition, the problem of parameter estimation has been essentially solved (Hambleton and Cook, 1977).

In his book entitled Probabilistic Models for Some Intelligence and Attainment Tests (1960), Georg Rasch, a Danish mathematician, introduced several mathematical models describing the test performance of students. One of these, which he called the simple item analysis model, has become popularly known as "The Rasch Model". Like all latent trait models, it supposes that test performance can be explained in terms of unobservable student characteristics, or traits. It is, however, the simplest of all the models, requiring only one item parameter.

The model defines the encounter of a student with an item in terms of the difference between the student's achievement, defined here as his position on some latent trait continuum, and the difficulty of the item scaled along the same continuum. Probabilistic statements about the success of the student on an item are then described by a one-parameter logistic function. Using notation similar to that of Wright (1977), the model can be written as:

$$\Pr\{x_{vi}/\beta_v, \delta_i\} = \frac{e^{x_{vi}(\beta_v - \delta_i)}}{1 + e^{(\beta_v - \delta_i)}}$$

where $\Pr\{x_{vi}/\beta_v, \delta_i\}$ represents the probability of outcome x_{vi} from the encounter between student "v" with achievement "β_v" and item "i" with difficulty "δ_i". $x_{vi} = 1$ if the student responds correctly; $x_{vi} = 0$ otherwise.

Other sources exist which delineate the Rasch model and its assumptions more fully (Hambleton and Cook, 1977; Wright, 1977; Rentz and Bashaw, 1975). The advantages of this approach over those of classical measurement theory, however, can best be understood in terms of the benefits derived from its application. Two basic outcomes are of interest, one involving the items mentioned above and the other involving the student. When properly applied, the model yields item difficulty estimates which are independent of the achievement levels of the students from which the estimates were developed, and student achievement estimates which are independent of the set of items from which those estimates were developed. Wright (1968) introduced the terms "sample-free item calibration" and "test-free person measurement" to describe these two desirable outcomes. The implications for educational measurement are item difficulty and student achievement estimates which

are invariant, a property referred to as "specifice objectivity" by Rasch.

The estimation of parameters is called "test calibration", and it involves obtaining two sets of information. One is difficulty estimates for each test item, while the other consists of a table of achievement estimates corresponding to each possible raw score on the test. These estimates are usually obtained by applying unconditional maximum likelihood estimation procedures to the test item response data (Wright and Douglas, 1977), and computer programs to accomplish this are available (for example, see Wright and Mead, 1977).

Rasch Calibrated Item Banks

The immediate relevance of the Rasch model to the solution of educational measurement problems may not as yet be fully apparent. However, when test items are constructed to fit the model, difficulty estimates obtained from a variety of different samples of students can be easily transferred onto a single common scale (Wright, 1977). Pools of commonly calibrated items can then be formed, referred to as "item banks", from which subsets can be drawn to make up tests. Since these subsets all share a common metric, test scores on all such tests are automatically equated, virtually solving the problem of test equating (Rentz and Bashaw, 1975, 1977).

The item bank concept has several important implications for the evaluation of student performance. Since achievement estimates obtained from any subset of bank items share a common metric regardless of which particular subset was used to determine those estimates, tests may be constructed from the bank which more accurately reflect the functional achievement level of the students to be measured. Even when each student tested responds to a different subset of bank items, comparable test information is available. Similarly, tests which more accurately reflect local curriculum

can be constructed from the bank, giving school district personnel more flexibility in the selection of achievement measures.

In order to utilize a Rasch calibrated item bank for Title I evaluation, however, some additional information is required. The relationship between the scale underlying the bank and a percentile scale based on national norms must be determined so that item bank scores can be converted into NCE units. Once this relationship is established, normative information becomes available for any subset of items drawn from the bank, provided that the subset is administered at an appropriate time during the school year. Since the normative information becomes associated with the item bank scale rather than with a particular subset of bank items, changes in the subset of items selected will not necessitate a redefinition of the relationship between the two scales. The advantages of Rasch calibrated item bank measurement, namely the opportunity to conduct functional level testing using instruments which reflect local curriculum, then become available for use in Title I evaluation.

Test Equating

One approach to the task of obtaining normative data for a Rasch calibrated item bank involves the use of classical test equating procedures. Test equating is the process of converting the system of units of one test to the system of units of the other so that scores derived from the two tests, after conversion, will be equivalent (Angoff, 1971). This process becomes meaningful only when both tests measure the same kind of achievement, when the conversion is unique except for a random error and is independent of the students tested (Angoff, 1971), and when both tests are equally reliable (Lord, 1977).

The task of providing normative information can be accomplished in the following manner. First, both a normed achievement test and an appropriate subset of items from the bank are administered to the same group of students. By applying test equating techniques to the pairs of test scores obtained, the system of units of one test is converted to the system of units of the other. As a result, each item bank raw test score has an equivalent "converted" normed test score. Since each normed test score also has a percentile rank associated with it, the normed data can be linked to the item bank raw scores. The item bank raw scores are then translated into bank achievement estimates via the item difficulty estimates already in existence in the bank for these items, and the link between item bank achievement estimates and normative data from which NCE units can be obtained is completed.

Several test equating techniques exist which are appropriate for use with pairs of test scores obtained from a single group of students. They can be generally classified as either linear and non-linear procedures, based on assumptions made about the relationship between the two score distributions of interest. When the shapes of these distributions are very similar, the results obtained from either class of techniques will be very similar. However, when this assumption appears questionable, a non-linear technique is generally preferable for ensuring equivalency of scores (Angoff, 1971). A non-linear technique based on the Rasch model was, therefore, used in the Project. This decision was further supported by the fact that Rasch model methodology was utilized in the development of the bank. A further consideration was the fact that this approach would result in converted scores in the lower range of the raw score distribution irrespective of the number of students actually obtaining these scores.

In order to investigate the adequacy of the Rasch Model equating technique employed, three additional procedures were used. A second more conventional non-linear technique, equipercentile equating, was applied to the test score pairs. In addition, two linear techniques, linear observed score equating and linear true score equating, were considered. A brief description of each of the four procedures follows.

Linear Observed Score Equating

Linear equating techniques are based on the assumption that the shapes of the raw score distributions of the tests to be equated are identical. If this assumption is tenable, the conversion of scores on one test to scores on the other can be accomplished by simply changing the origin and unit of measurement of one of the distributions (i.e., by applying a linear transformation to the scores). A more formal definition of equivalent scores obtained from applying a linear equating technique states that scores on two tests are equivalent if they correspond to equal standard-score deviates (Angoff, 1971).

For a given raw score on test X, denoted x here, a score x' can be obtained by applying the following transformation:

$$x' = Ax + B,$$

where $A = s_y/s_x$ and $B = \bar{y} - A\bar{x}$. Here s_x and s_y are the standard deviations of the X and Y raw score distributions, respectively, and $\bar{x}$ and $\bar{y}$ represent the means respectively. The distribution of converted scores (X') obtained in this fashion is identical to that of the unconverted scores (Y) (i.e., the mean and standard deviation of the Y distribution). If the X and Y distributions have proportionally the same shape, then the X' and Y

distributions will be identical, and the scores obtained will, therefore, be comparable.

Like all statistical procedures, equating is subject to random error due to sampling fluctuations of the means and standard deviations of scores on X and Y. The standard error of equating by this method is given by:

$$SE_{x} = \sqrt{s_x^2 (1-r_{xy}) \left[\frac{Z_y^2 (1+r_{xy}) + 2}{N} \right]}, \qquad (2)$$

where N is the number of students taking the test pair, r_{xy} is the correlation between scores on X and Y, and $Z_y = (y-\bar{y})/s_y$ (Lord, 1950).

For the purposes of the present application, test X represents an appropriate subset of items drawn from the item bank, while test Y represents the corresponding CAT/C test, both test pair members being administered to the same group of students.

Linear True Score Equating

When the assumption of equal test reliability across equating test pairs is a concern, a second method of linear equating, involving estimated true scores rather than observed scores, is preferable. The equation for transforming scores on test X to the scale of test Y is of the same form as (1) above, except that estimates of the true score standard deviations, s_{tx} and s_{ty} are substituted for s_x and s_y, respectively. The substituted values are given by:

$$s_{tx} = \sqrt{s_x^2 - SEM_x^2} \qquad \text{and} \qquad (3)$$

$$s_{ty} = \sqrt{s_y^2 - SEM_y^2},$$

where SEM_x and SEM_y are the standard errors of measurement for tests X and Y, respectively.

Equipercentile Equating

Equipercentile equating is the non-linear equating technique most widely used. It is generally applied when the assumption of identical score distributions mentioned earlier is questionable, and it involves stretching and compressing the scale of one test so that its distribution coincides with that of the other. A commonly accepted definition of equivalent scores obtained from applying an equipercentile procedure states that two scores may be considered equivalent if their corresponding precentile ranks in any given group are equal (Angoff, 1971).

Equipercentile equating is accomplished in the following manner. Let x_i and y_m represent scores i and m on tests X and Y respectively, and P_{x_m} and P_{y_m} represent their respective percentile ranks. Then x_i, the converted x score, is determined such that $P_{y_m} < P_{x_i} < P_{y_{m+1}}$. By using linear interpolation, x_i' is obtained as follows:

$$x_i' = \frac{\left(P_{x_i} - P_{y_m}\right)\left(y_{m+1} - y_m\right)}{\left(P_{y_{m+1}} - P_{y_m}\right)} + y_m$$

When the percentile rank of score x_i is exactly equal to a number of consecutive percentile ranks of y scores, the converted score x_i' is equal to the average score represented by the equal percentile ranks:

$$x_i' = \left(\frac{y_{m+k} - y_m}{2}\right) + y_m ,$$

where k is equal to one less than the number of y scores with the same percentile rank. If $k = 0$, that is, if the percentile rank of score x_i

is equal to only one y percentile rank, then x_i is equal to y_m (Wichert, 1976).

Rasch Model Equating

The last equating procedure considered involves the use of parameter estimates obtained from a Rasch analysis of the two tests to be equated. Here the definition of equivalent scores states that scores obtained from two tests are said to be equivalent if they correspond to the same Rasch achievement estimate (Rentz and Bashaw, 1975).

Initially, item difficulty estimates for all items on both tests are obtained through a single analysis. The difficulty estimates for only those items appearing on the first test are then used to generate a raw score to achievement estimate table for that test. The following estimation equation is used:

$$r = \sum_{i=1}^{k} \left(e^{(b_r - d_i)} / \left(1 + e^{(b_r - d_i)} \right) \right),$$

where "b_r" represents the ability estimate for raw score "r", and "d_i" represents the item difficulty estimate for item "i" appearing on the first test obtained from the initial analysis of both tests combined. Iterative estimation procedures for solving for b_r, given r and the d_i's, can be found in Wright and Stone (1979, pages 142-143).

Next, given the b_r values obtained for the first test and the d_i values for items appearing on the second test, a conversion table for the second test can be generated. Raw score to raw score conversions can be obtained for the two tests in this manner.

Model Data Fit

While the Rasch model has important theoretical and practical implications for the solution of measurement problems, it also involves some rather strong assumptions about the nature of the data under consideration. For this reason, the question of whether a particular set of data "fit" the model is an important concern. Lack of fit poses a threat to the validity of any results obtained from an application of the model. The Rasch model entails the strongest assumptions of all the latent trait models, further increasing the need for assessing fit.

The question of model data fit is a complex one. The distinction made by Rentz and Bashaw (1975) between Rasch model assumptions and those conditions necessary for fit is useful here. Specifically, they suggest that unidimensionality of the latent trait, equal item discrimination and guessing can best be thought of as conditions necessary for fit rather than as model assumptions. This conceptualization permits the more formal statement of model assumptions (Rasch, 1966, page 50) to be translated into operational constructs which are data related.

Several methods for determining model data fit have been suggested (Andersen, 1973; Wright and Panchapaksen, 1969). However, the validity of these chi-square approaches appears questionable given the asymptotic nature of the test statistic distribution (Hambleton et al., 1978). A recent approach which seems promising involves fitting a least squares line to the item responsed data within the range where the data are approximately linear (George, 1979). A t-test is then used to determine the probability that the observed data were sampled from a population in which the slope of the fitted line is the same as the slope of the theoretical

Rasch item characteristic curve within that range. The test statistic is computed as follows:

$$t = \left(\frac{\hat{\beta}_i - .20940}{s}\right)\left(\sqrt{SS_x}\right),$$

where:

$$SS_x = \sum_{v=1}^{N} (b_v - d)^2 - \frac{\left(\sum_{v=1}^{N} (b_v - d)\right)^2}{N},$$

$$SS_y = \sum_{v=1}^{N} x_v^2 - \frac{\left(\sum_{v=1}^{N} x_v\right)^2}{N},$$

$$SS_{xy} = \sum_{v=1}^{N} (b_v - d)x_v - \frac{\left(\sum_{v=1}^{N} (b_v - d)\right)\left(\sum_{v=1}^{N} x_v\right)}{N},$$

$$\hat{\beta}_i = SS_{xy}/SS_x,$$

$$SSE = SS_y - (SS_{xy})^2/SS_x$$

and

$$s = \sqrt{\frac{SSE}{N-2}}.$$

Here "b_v" is the achievement estimate for student "v", "d" is the item difficulty estimate for the item being assessed, and $x_v = 1$ if the student "v" answered the item correctly; otherwise $x_v = 0$. This test statistic follows a t-distribution, with (N-2) degrees of freedom, and the least squares lines is fitted between $\pm 2.0(b - d)$.

This particular test statistic will shed some light on the question of equal item discrimination values, one of the data conditions necessary for fit. Model data fit is, however, a much larger concern involving a variety of different sources of deviation. Unfortunately, at the present time no commonly accepted evaluation procedures have been identified. In short, assessment of model data fit needs considerable further research (Hambleton et al., 1978).

METHODOLOGY

Sample

Over 6,000 students were involved in the Project, about 1,000 at each grade level. The Project sample was obtained by inviting schools from two Oregon school districts to participate. Twenty-three schools accepted the invitation, and the sample consisted of all students enrolled in grades three through eight in these schools. Table 1 presents the exact number of students tested at each grade level for each equating test pair.

Instrumentation

The Northwest Evaluation Association (NWEA), a consortium of school districts from Oregon and Washington states, was founded in 1973 for the purpose of developing goal-referenced item banks in various school subjects (Northwest Evaluation Association, 1978). At present, two Rasch calibrated item banks in the areas of reading and mathematics have been developed which span grades three through eight. Both banks contain over 1,100 items. All items are referenced to their respective collection of course and program goals developed by the Tri-County Course Goals Project and have been field-tested on relatively large samples of students before inclusion in the banks. Both banks were utilized in the Title I Linking Project.

TABLE 1

Number of Students Tested by Grade and Test Equating Pair

READING

GRADE	CAT/C TEST CODE	ITEM BANK TEST CODE	NUMBER TESTED
3	CAT13	318	488
4	CAT14	319	515
5	CAT15	320	500
6	CAT16	320	489
7	CAT17	321	533
8	CAT18	321	601
		TOTAL	3,086

MATHEMATICS

GRADE	CAT/C TEST CODE	ITEM BANK TEST CODE	NUMBER TESTED
3	CAT13	818	446
4	CAT14	819	439
5	CAT15	820	481
6	CAT16	820	489
7	CAT17	821	612
8	CAT18	821	612
		TOTAL	3,079

Four item bank tests in each of the two content areas of interest were selected for administration. These instruments were assembled from the NWEA item banks by the Portland School District 1J for use in their district-wide evaluation program (Forster and Doherty, 1978). Each instrument was originally part of a series of overlapping tests designed to measure reading or mathematics achievement in grades three through eight. Utilizing data on test difficulty obtained by the Portland School District in their fall 1978 testing program, only four tests were selected from each series for use in the Project. These tests are presented in Table 2 by grade level along with their corresponding test codes and the difficulty ranges within which test items were drawn from the item banks.

An informal survey of several education service districts in Oregon indicated that the California Achievement Tests, Form C (CAT/C), CTB/McGraw-Hill publishers, and the Iowa Tests of Basic Skills (ITBS), Houghton-Mifflin publishers, were two commonly used nationally normed achievement tests in Oregon. Since both instruments appeared to be equally appropriate for use by the Project, final selection was based on the availability of an adequate sample. Accordingly, the CAT/C was selected.

The CAT/C measures achievement in the areas of prereading, reading, spelling, language, mathematics, and reference skills. The battery consists of ten tests which overlay in difficulty, spanning grades K-12 (California Achievement Tests, 1979). Only those normed for

TABLE 2

CAT/C Test Code and Grade Coverage Along With the Corresponding Item Bank Test Code and Difficulty Range by Grade

			READING		MATHEMATICS	
GRADE	CAT/C TEST CODE	CAT/C GRADE COVERAGE	ITEM BANK TEST CODE	ITEM BANK DIFFICULTY RANGE IN RIT[a] UNITS	ITEM BANK TEST CODE	ITEM BANK DIFFICULTY RANGE IN RIT[a] UNITS
3	CAT13	2.5 - 3.9	318	145 - 219	818	141 - 224
4	CAT14	3.5 - 4.9	319	156 - 230	819	152 - 234
5	CAT15	4.5 - 5.9	320	167 - 239	820	165 - 247
6	CAT16	5.5 - 6.9	320	167 - 239	820	165 - 247
7	CAT17	6.5 - 7.9	321	180 - 251	821	179 - 259
8	CAT18	7.5 - 9.9	321	180 - 251	821	179 - 259

[a]The Rasch units scale (RIT) was originally introduced by the Northwest Evaluation Association to define the underlying item bank scales. Scale values range from about 150 to 250 RIT's, with the average fifth grade value being set at 200. The linear transformation used to obtain these scale values is as follows:

$$RIT = 10 \times (\text{item difficulty estimate in logits}) + 200$$

administration at grades three through eight, coded CAT13 through CAT18, were considered. Table 2 also presents the selected CAT/C tests utilized along with the grade coverage of each test.

Data Collection

All students were administered the CAT/C test appropriate for their grade placement. Functional level testing was not conducted since this portion of the data collection was part of the routine assessment efforts of the participating school districts and was, therefore, beyond the author's control. In addition to the CAT/C, approximately half of the students at each grade level was administered an appropriate reading item bank test, while the remaining half took an appropriate mathematics item bank test. Decisions as to whether a school would participate in the reading or mathematics portion of the item bank testing were based on their preference when expressed or on random assignment.

All testing was supervised within each school by the school test coordinator, although the actual test administration was conducted by individual classroom teachers. All data were collected during late March and early April of 1979. While the exact test dates varied among schools, all data were collected within the same week in a particular school. The appropriate CAT/C test was administered first, followed by the corresponding item bank test. Where possible, make-up tests were given to those students absent on any of the test dates.

Conversion Table Development

The development of tables which permit the conversion of item bank achievement estimates expressed in RIT units into their NCE equivalents involved several steps. As discussed earlier, the application of Rasch model equating techniques to the pairs of item bank and CAT/C raw scores yielded a CAT/C converted score for each item bank raw score. (See Appendix B for the values obtained at each grade in each content area.) First, the item bank raw scores were converted into item bank achievement estimates expressed in RIT units. This was accomplished by applying the UCON procedures outlined by Wright and Stone (1979, pages 142-143) to the item difficulty estimates already recorded in the bank for these items. Next, the percentile ranks associated with the CAT/C converted scores were determined. These values were obtained through interpolation from the raw score to percentile rank tables provided by the CAT/C publishers (California Achievement Tests, 1978). Values for both the beginning and end of each grade were determined. Last, these percentile ranks were converted into NCE units, thus completing the link from RIT achievement estimates to NCE units.

RESULTS

Data Screening

Before the analysis phase of the Project was begun, the data collected were screened in two ways. The first involved an examination of student item responses by equating test pairs. Any student who had not attempted both tests in any test pair was excluded, as well an any student who had not responded to at least 35 percent of the items on either of the tests in a given pair. This approach was selected in an attempt to avoid biasing the equating results by the inclusion of data from students for whom the test pairs were inappropriate. Table 3 indicates the number of students responding to each test pair after these deletions, as well as the percent of the original sample this number represents.

A second kind of data screening was also performed, and it involved an examination of the stability of bank item difficulty estimates. While no CAT/C test items could be deleted since the norms used in the Project were developed for total CAT/C test scores, some selectivity could be exercised with regard to items appearing on the bank tests. In order to identify unstable items, each bank test was calibrated using the Rasch item analysis program (BICAL) developed by Wright and Mead (1977). The difficulty estimates obtained were then rescaled by multiplying each estimate by ten and adding it to 200. The average difference between these rescaled estimates and those recorded in the bank for the same items was then computed.

TABLE 3

Number of Students Remaining After Initial Data Screening with Percent of Original Sample Retained in Parentheses

GRADE	READING	MATHEMATICS
3	463 (95%)	343 (77%)
4	457 (88%)	354 (81%)
5	481 (96%)	370 (77%)
6	462 (94%)	429 (88%)
7	479 (90%)	549 (90%)
8	559 (93%)	559 (91%)
TOTAL	2,898 (94%)	2,604 (85%)

TABLE 4

The Linking Constants Added At Each Grade to the Rescaled Item Difficulty Estimates Along with the Standard Deviation of the Difference Between These Estimates and Those Recorded in the Bank

GRADE	READING	MATHEMATICS
3	-17.2	-17.5
4	- 5.5	- 5.9
5	1.8	7.7
6	1.8	7.7
7	17.9	20.6
8	17.9	20.6
Standard Deviation of the Difference	3.4	4.4

This value, known as the "linking constant," was then added to each rescaled estimate. Theoretically, the difference between these final estimates and those recorded in the bank should now be zero, except for random measurement error.

In order to determine how different these two sets of estimates actually were, a standard deviation of the differences was computed for each content area across all grades. These standard deviations, along with the values of the linking constants added to each rescaled estimate, can be found in Table 4. Any bank item whose pair of difficulty estimates differed by two or more standard deviation units was deleted from further analyses. The number of items deleted from each bank test along with the total number retained can be seen in Table 5.

Total Test Summary Statistics

After initial data screening, summary statistics based on those students and items retained were compiled for each test. Total test raw score means, standard deviations and KR-20 reliability estimates are reported in Table 6. Correlations between raw scores for each equating test pair, along with the values of these correlations corrected for attenuation, appear in Table 7.

Although only total test score information was utilized in the Project, both the item bank and CAT/C tests were composed of

(continued on page 28)

TABLE 5

Number of Unstable Items Deleted From Each Item Bank Test Along with the Total Number Retained by Equating Test Pair

READING

GRADE	NUMBER OF BANK ITEMS DELETED	NUMBER OF BANK ITEMS RETAINED	NUMBER OF CAT/C ITEMS
3	1	44	73
4	3	42	70
5	0	45	70
6	3	42	70
7	2	43	70
8	1	44	70

MATHEMATICS

GRADE	NUMBER OF BANK ITEMS DELETED	NUMBER OF BANK ITEMS RETAINED	NUMBER OF CAT/C ITEMS
3	1	54	85
4	2	53	85
5	3	52	85
7	7	48	85
1	1	49	85
8	3	47	85

TABLE 6

Total Test Raw Score Means, Standard Deviations and Reliability Estimates by Grade

READING

	ITEM BANK TEST				CAT/C TOTAL TEST		
GRADE	N	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20
3	463	31.1	9.0	.92	59.2[a]	11.0[a]	.93[a]
4	457	27.4	7.6	.88	49.6	13.5	.94
5	481	28.8	8.6	.89	51.7	12.1	.93
6	462	28.7	8.2	.90	48.3	12.5	.93
7	479	22.9	7.5	.84	46.8	12.3	.92
8	559	26.2	8.1	.87	43.8	13.0	.93

MATHEMATICS

	ITEM BANK TEST				CAT/C TOTAL TEST		
GRADE	N	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20
3	343	36.9	10.1	.91	56.5	14.6	.94
4	354	34.8	11.0	.93	52.0	15.9	.94
5	370	29.0	10.0	.90	51.9	15.5	.94
6	429	33.1	8.9	.90	52.2	14.6	.93
7	549	29.1	10.1	.91	48.6	16.5	.94
8	559	31.5	9.6	.91	55.0	16.9	.95

[a]Total test information for CAT13 is based on all 73 items appearing on the reading test. This includes the 20 Phonic Analysis items as well as the 11 Structural Analysis items for a total of 73 items.

TABLE 7

Item Bank and CAT/C Total Test Raw Score Correlations Uncorrected and Corrected for Attenuation by Content Area and Grade

	READING		MATHEMATICS	
Grade	r	r corrected for attenuation	r	r corrected for attenuation
3	.70	.76	.87	.94
4	.82	.90	.87	.93
5	.83	.91	.85	.92
6	.85	.93	.88	.96
7	.79	.90	.88	.94
8	.84	.93	.84	.90

subtests. They were vocabulary and comprehension or computations and concepts for reading and mathematics, respectively. For the sake of completeness, summary subtest information has been included in Appendix A. Tables 1 and 2 present the number of items and percent of the total this number represents in each subtest by equating pair and grade. Raw score means, standard deviations and KR-20 reliability estimates are given in Tables 3 and 4 of Appendix A. Table 5, also in Appendix A, presents the within test subtest correlations, while Table 6 presents the between test subtest correlations. Both tables include the value of the correlations corrected for attenuation.

Model Data Fit

The slope of the item characteristic curve for each test item was compared to that of the theoretical Rasch model curve by fitting a least squares line to the observed data within the range $\pm 2.0(b - d)$ (George, 1979). All items, even those excluded after the initial data

screening were evaluated in this manner. The item slopes were compared using a t-test, and the percent found to differ significantly was determined using the three different probability levels of significance ($p<.01$, $p<.05$, $p<.10$). This information can be seen in Table 8.

Test Equating

Four equating procedures were applied to each set of raw test score pairs. Two linear methods, one based on observed scores and one on estimated true scores, as well as two nonlinear methods, equipercentile equating and a Rasch model technique, were utilized.

The converted CAT/C scores yielded by each procedure can be found in Appendix B, Tables 1 through 12. Tables 1 through 6 present the converted CAT/C total reading scores obtained for each item bank raw score for grades three through eight, respectively. A frequency distribution of item bank raw scores is also included. Tables 7 through 12 present the converted CAT/C total mathematics scores obtained.

A graphic representation of Appendix B can be found in Appendix C, Figures 1 through 12. It should be noted that while all four equating techniques yield a converted CAT/C score for every possible item bank raw score, not all item bank raw scores were actually observed in the Project sample. This is especially true at the lower end of the raw score distribution. For the sake of completeness, all raw score pairs have been included in both Appendix B and C.

TABLE 8

Percent of Test Items with Item Characteristic Curve Slopes Which Differ Significantly from the Slope of the Theoretical Rasch Model Curve Using Three Different Probability Levels as Criteria[a]

READING

	ITEM BANK				CAT/C			
GRADE	TOTAL ITEMS	p<.01	p<.05	p<.10	TOTAL ITEMS	p<.01	p<.05	p<.10
3	45	29%	38%	44%	63	14%	25%	35%
4	45	31%	40%	44%	70	17%	27%	36%
5	45	29%	44%	53%	68	16%	31%	40%
6	45	24%	40%	47%	68	22%	32%	43%
7	45	22%	38%	44%	69	19%	32%	41%
8	45	33%	44%	53%	69	35%	54%	65%

MATHEMATICS

	ITEM BANK				CAT/C			
GRADE	TOTAL ITEMS	p<.01	p<.05	p<.10	TOTAL ITEMS	p<.01	p<.05	p<.10
3	54	9%	19%	31%	85	11%	19%	25%
4	55	11%	25%	35%	84	11%	24%	28%
5	55	20%	29%	38%	85	19%	25%	29%
6	55	18%	31%	40%	85	19%	27%	33%
7	50	26%	40%	42%	85	22%	37%	45%
8	50	18%	32%	34%	85	29%	39%	48%

[a]Percentages were computed using only those items with at least 50 observations between $\pm 2.0(b - d)$.

Equating Comparisons

In order to make comparisons among the results yielded by each of the four procedures, an average absolute discrepancy was computed for each item bank raw score. This value was obtained by first summing the absolute differences between the CAT/C converted score yielded by each procedure and those actually observed at each item bank raw score. This sum was then divided by the number of differences summed. The values obtained can be found in Appendix D, Tables 1 through 12.

Next, an overall average absolute discrepancy was computed at each grade for each technique. This value was obtained by summing absolute differences across all item bank raw scores, and then dividing the sum obtained by the total number of differences summed. The overall average absolute discrepancy computed for each equating technique is presented by grade in Table 9.

In order to determine whether the overall discrepancies produced by the four procedures differed significantly, a one-way repeated measures analysis of variance was performed at each grade using the absolute discrepancies observed as the dependent variable. The results of these analyses are summarized in Table 10 along with the proportion of variance accounted for by the equating methods.

Rasch Model Equating Bias

While the average absolute discrepancy between CAT/C converted and observed scores is useful in assessing the accuracy of the equating results obtained, the distribution of these discrepancies rather than their average absolute value might be helpful in evaluating equating bias. Graphs of these values

TABLE 9

Overall Average Absolute Discrepancy Between CAT/C Converted Scores Yielded by Four Equating Methods and Those Actually Observed at Each Grade for Each Content Area (LO--Linear Observed; LT--Linear True; EQ--Equipercentile; RM--Rasch Model)

READING

GRADE	LO	LT	EQ	RM
3	6.02	6.04	6.05	6.05
4	6.24	6.37	5.97	5.89
5	5.40	5.48	5.24	5.08
6	5.26	5.31	5.27	5.29
7	6.17	6.34	6.12	5.90
8	5.68	5.77	5.67	5.53

MATHEMATICS

GRADE	LO	LT	EQ	RM
3	5.71	5.75	5.68	5.78
4	6.45	6.47	6.30	6.48
5	6.55	6.63	6.46	6.52
6	5.81	5.85	5.77	6.10
7	6.41	6.47	6.16	6.32
8	6.89	6.96	6.96	6.84

TABLE 10

Summary of the One-Way Repeated Measures Analyses of Variance Conducted at Each Grade Level Using the Absolute Discrepancy Between Converted and Observed CAT/C Scores as the Dependent Variable

READING

Grade	F	df Error[1]	Probability	Proportion of Variance[2]
3	.16	1365	.92	.0000
4	25.42	1368	< .01	.0013
5	16.11	1434	< .01	.0010
6	1.30	1377	.27	.0000
7	10.58	1434	< .01	.0010
8	14.76	1674	< .01	.0003

MATHEMATICS

Grade	F	df Error[1]	Probability	Proportion of Variance[2]
3	2.60	1020	.05	.0001
4	2.10	1059	.10	.0002
5	4.55	1107	< .01	.0001
6	7.75	1281	< .01	.0009
7	11.00	1638	< .01	.0005
8	5.57	1659	< .01	.0001

[1]For all F-values reported, the treatment degrees of freedom are 3.

[2]Proportion of variance accounted for was computed as the ratio of treatment to total sum of squares.

for the Rasch model equating technique can be found in Appendix E, Figures 1 through 12. Pearson product-moment correlations between these discrepancies and their corresponding item bank raw scores were computed. These values can be seen in Table 11 for both reading and mathematics. All correlations, with the exception of the one reported for grade 7 reading, were significantly different from zero at the .05 level.

TABLE 11

Pearson Correlations between CAT/C Converted and Observed Score Discrepancies and Their Corresponding Item Bank Raw Scores for the Rasch Model Equating Technique

Grade	Reading	Mathematics
3	-.39	-.22
4	-.14	-.27
5	-.12	-.25
6	-.24	-.36
7	-.08	-.21
8	-.14	-.27

RIT to NCE Conversion Tables

One of the Project objectives was the development of tables which facilitate the conversion of item bank achievement estimates expressed in RIT units into their NCE equivalents. The procedure followed in this development was outlined earlier. The RIT to NCE conversion tables can be found in Tables 12 through 17 for both reading and mathematics.

An Application

For a number of years Portland School District 1J has been using tests developed from Rasch calibrated item banks for routine district-wide evaluation. In the fall of 1978, and again in the spring of 1979,

(continued on page 44)

TABLE 12

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 3 and 4, Reading

GRADE 3			GRADE 4		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
225.18	99.00	90.77	236.84	99.00	98.48
217.76	94.50	84.40	229.32	99.00	91.86
213.21	91.00	79.00	224.69	96.49	85.25
209.83	86.68	74.64	221.24	89.10	79.92
207.09	81.89	70.54	218.44	84.86	75.19
204.76	78.30	67.06	216.05	80.84	71.28
202.70	75.81	64.21	213.94	78.07	67.92
200.85	73.09	61.24	212.03	74.84	65.61
199.15	70.24	58.82	210.28	72.36	62.99
197.56	67.66	56.96	208.66	70.29	61.21
196.07	65.86	54.68	207.13	68.72	59.43
194.66	63.67	53.24	205.67	67.08	57.69
193.31	62.09	51.45	204.28	65.27	56.21
192.01	60.12	49.74	202.94	63.56	54.81
190.76	58.78	48.49	201.65	62.17	53.65
189.54	57.39	46.87	200.39	60.86	52.46
188.35	56.03	45.35	199.15	58.97	51.17
187.18	54.67	43.98	197.94	57.10	49.57
186.03	53.30	42.79	196.74	55.82	48.36
184.90	51.93	41.69	195.56	54.03	46.99
183.78	50.56	40.20	194.38	52.73	45.52
182.66	49.16	38.77	193.20	50.95	44.06
181.54	47.74	37.57	192.01	49.15	42.66
180.42	46.81	36.44	190.82	47.86	41.26
179.30	45.29	34.84	189.62	46.53	40.02
178.16	43.72	33.82	188.40	44.65	38.22
177.01	42.65	32.77	187.15	43.24	37.26
175.84	41.40	31.65	185.87	41.21	35.42
174.64	39.92	29.92	184.55	38.72	33.45
173.42	38.28	28.30	183.19	36.90	32.03
172.15	37.05	27.67	181.76	33.71	29.08
170.84	35.47	26.05	180.26	30.16	25.56
169.46	33.72	24.21	178.67	25.87	22.48
168.01	31.84	23.26	176.95	20.18	16.88
166.47	29.61	20.88	175.09	15.47	11.79
164.82	27.58	19.39	173.02	9.01	6.67
163.01	23.60	15.88	170.67	5.26	1.00
161.01	20.45	13.22	167.91	1.00	1.00
158.73	14.54	9.05	164.51	1.00	1.00
156.04	6.67	1.00	159.92	1.00	1.00
152.71	1.00	1.00	152.44	1.00	1.00
148.21	1.00	1.00			
140.84	1.00	1.00			

TABLE 13

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 5 and 6, Reading

GRADE 5			GRADE 6		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
243.04	99.00	99.00	242.47	99.00	99.00
235.66	99.00	99.00	235.05	94.16	90.44
231.16	99.00	90.37	230.52	89.45	82.67
227.85	93.70	85.06	227.19	83.70	77.65
225.20	87.34	81.43	224.50	79.71	73.00
222.95	84.75	78.44	222.23	75.75	69.64
220.99	81.24	74.89	220.24	72.51	66.03
219.23	79.50	71.75	218.45	69.79	63.28
217.63	76.76	69.16	216.82	67.38	60.77
216.14	74.35	67.30	215.31	65.19	58.52
214.76	72.18	64.98	213.89	62.91	56.68
213.45	70.20	62.96	212.55	60.68	55.00
212.20	68.38	61.22	211.27	58.75	52.83
211.01	66.13	59.57	210.04	56.66	51.26
209.85	64.92	58.20	208.85	54.83	49.68
208.73	63.35	56.29	207.69	53.28	48.10
207.64	61.59	54.93	206.55	51.26	46.50
206.57	59.70	53.59	205.44	49.68	45.32
205.52	58.77	52.21	204.34	48.10	43.77
204.48	57.33	50.84	203.24	46.48	42.55
203.44	55.86	49.44	202.15	44.85	41.10
202.41	54.38	48.02	201.05	43.63	39.40
201.38	52.87	47.08	199.96	42.01	38.31
200.35	51.38	45.90	198.84	40.24	36.55
199.31	49.86	44.52	197.72	38.62	35.47
198.26	48.30	42.88	196.57	37.05	33.63
197.19	46.69	41.73	195.39	34.94	31.80
196.10	45.38	40.35	194.18	33.37	29.71
194.99	43.81	38.51	192.93	30.91	27.53
193.84	42.00	37.25	191.62	28.20	24.71
192.66	40.40	36.08	190.26	24.92	21.58
191.43	38.39	34.21	188.81	21.98	17.92
190.15	36.37	32.90	187.26	18.12	14.20
188.80	33.86	30.77	185.60	12.51	10.33
187.37	31.20	28.38	183.77	6.67	6.65
185.84	28.66	25.52	181.74	2.49	1.00
184.19	24.69	22.05	179.43	1.00	1.00
182.38	19.32	17.49	176.71	1.00	1.00
180.37	13.12	13.12	173.34	1.00	1.00
178.07	9.27	6.67	168.78	1.00	1.00
175.35	1.00	1.00	161.34	1.00	1.00
172.00	1.00	1.00			
167.45	1.00	1.00			
160.03	1.00	1.00			

TABLE 14

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 7 and 8, Reading

GRADE 7			GRADE 8		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
256.72	99.00	99.00	257.17	99.00	99.00
249.43	99.00	93.97	249.89	99.00	99.00
245.00	99.00	93.33	245.47	99.00	95.61
241.75	94.45	88.45	242.23	93.48	89.81
239.15	90.69	85.27	239.64	89.67	83.07
236.94	87.67	82.01	237.44	87.01	79.72
235.02	83.75	79.96	235.53	83.29	76.09
233.30	82.80	77.57	233.82	79.82	73.05
231.72	81.37	75.11	232.26	77.55	70.45
230.27	79.66	72.84	230.81	75.06	68.11
228.91	76.93	71.72	229.46	72.55	66.73
227.63	75.61	69.78	228.19	70.68	64.68
226.40	73.99	67.92	226.98	68.42	62.82
225.23	72.24	66.62	225.81	66.54	61.27
224.09	71.14	65.22	224.69	64.87	59.83
222.99	69.34	63.56	223.60	62.93	58.17
221.91	67.54	62.58	222.54	61.16	56.54
220.86	66.25	61.03	221.50	59.42	54.92
219.82	64.97	59.92	220.48	57.74	53.35
218.79	63.29	58.60	219.47	56.11	51.79
217.77	61.59	57.50	218.46	54.09	50.24
216.75	60.47	55.94	217.46	52.36	48.68
215.73	58.83	54.38	216.46	50.80	47.12
214.70	57.15	53.24	215.46	48.70	45.09
213.66	55.83	51.72	214.45	47.12	43.85
212.61	54.31	50.57	213.43	45.48	41.74
211.53	52.62	48.68	212.39	43.39	40.00
210.43	50.90	47.22	211.33	41.53	38.12
209.30	49.13	45.40	210.24	38.79	36.11
208.13	47.31	44.05	209.12	36.99	33.85
206.91	45.37	42.05	207.96	34.45	31.52
205.63	43.32	39.83	206.75	31.37	28.95
204.27	40.57	37.50	205.48	28.55	25.90
202.83	38.07	34.23	204.13	24.52	22.15
201.26	34.64	31.11	202.69	20.41	18.83
199.55	30.47	27.40	201.14	14.99	14.99
197.64	24.72	22.40	199.43	10.54	10.54
195.44	16.53	16.85	197.53	2.96	2.96
192.85	7.49	11.00	195.34	1.00	1.00
189.62	1.00	1.00	192.75	1.00	1.00
185.22	1.00	1.00	189.52	1.00	1.00
177.93	1.00	1.00	185.12	1.00	1.00
			177.85	1.00	1.00

TABLE 15

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 3 and 4, Mathematics

GRADE 3			GRADE 4		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
226.42	99.00	97.49	236.66	99.00	99.00
219.11	99.00	88.83	229.35	99.00	99.00
214.67	99.00	82.37	224.92	99.00	92.02
211.41	99.00	78.06	221.68	99.00	85.80
208.79	99.00	73.97	219.09	99.00	80.22
206.59	95.02	70.76	216.90	95.12	76.28
204.65	93.33	68.04	215.00	91.29	72.71
202.93	89.94	65.69	213.31	86.80	69.79
201.36	88.19	63.85	211.78	84.24	67.46
199.92	84.60	61.53	210.36	81.23	64.95
198.56	83.41	60.03	209.05	78.69	63.10
197.29	81.14	58.60	207.82	76.21	60.67
196.08	80.00	57.33	206.66	73.75	59.30
194.93	78.00	55.67	205.55	72.22	57.25
193.83	76.14	54.07	204.49	69.87	55.19
192.75	73.54	52.46	203.47	67.95	53.90
191.72	72.15	50.91	202.49	66.01	52.00
190.72	70.01	49.38	201.53	63.94	50.84
189.74	68.65	47.85	200.60	62.02	48.85
188.77	67.00	46.84	199.69	60.25	47.74
187.83	65.66	45.27	198.80	58.58	46.22
186.90	63.69	43.68	197.92	56.95	45.17
185.98	61.86	42.64	197.06	55.37	43.59
185.08	60.59	41.44	196.20	53.85	42.57
184.18	59.06	39.87	195.35	52.32	41.41
183.28	57.41	38.37	194.51	50.84	39.85
182.39	56.38	37.67	193.67	49.37	38.39
181.50	54.79	36.14	192.83	47.91	37.76
180.61	53.76	35.18	191.98	46.41	36.22
179.71	51.72	33.45	191.13	44.93	34.77
178.81	50.63	32.43	190.28	43.39	33.61
177.90	49.10	31.30	189.41	41.85	32.62
176.99	47.54	30.17	188.54	40.27	30.83
176.06	46.06	28.98	187.64	38.12	29.67
175.11	44.03	27.66	186.73	36.43	27.85
174.14	42.41	26.28	185.80	34.32	26.22
173.16	39.81	24.25	184.84	32.33	24.26
172.14	37.83	21.83	183.86	29.40	21.91
171.10	35.65	21.24	182.83	26.73	20.02
170.01	32.53	19.02	181.76	23.75	17.80
168.88	30.12	16.55	180.64	20.24	15.14
167.70	27.51	13.46	179.45	16.93	11.57
166.45	23.77	12.06	178.19	13.15	6.67
165.13	20.68	6.96	176.83	8.69	4.53

RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
163.71	17.37	4.71	175.35	6.67	1.00
162.17	13.15	1.00	173.73	1.00	1.00
160.49	8.28	1.00	171.90	1.00	1.00
158.59	1.00	1.00	169.79	1.00	1.00
156.41	1.00	1.00	167.29	1.00	1.00
153.84	1.00	1.00	164.14	1.00	1.00
150.62	1.00	1.00	159.81	1.00	1.00
146.22	1.00	1.00	152.62	1.00	1.00
138.95	1.00	1.00			

TABLE 16

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 5 and 6, Mathematics

GRADE 5			GRADE 6		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
252.98	99.00	99.00	251.40	99.00	99.00
245.35	99.00	99.00	243.78	99.00	93.70
240.63	99.00	99.00	239.07	99.00	88.06
237.11	99.00	98.90	235.55	99.00	82.56
234.26	99.00	91.84	232.70	93.33	78.81
231.84	99.00	87.63	230.27	89.67	74.90
229.73	99.00	82.62	228.15	84.60	71.69
227.83	95.52	80.34	226.25	82.19	68.83
226.11	93.33	77.24	224.51	78.99	66.42
224.52	89.67	74.05	222.92	76.44	63.90
223.04	87.45	72.03	221.43	73.26	61.72
221.66	84.18	69.76	220.03	70.28	58.85
220.35	81.63	67.96	218.71	67.52	57.25
219.10	79.39	66.11	217.45	64.69	54.65
217.91	76.76	64.07	216.24	62.31	52.68
216.76	74.69	62.78	215.08	60.14	50.80
215.65	73.13	60.98	213.96	57.84	48.95
214.58	70.81	59.21	212.86	55.51	47.13
213.54	68.86	57.80	211.79	53.24	45.28
212.52	67.60	56.35	210.74	50.93	43.42
211.51	65.42	54.68	209.71	49.17	42.13
210.53	63.44	53.09	208.69	46.95	40.25
209.56	62.02	51.56	207.69	45.15	38.57
208.60	60.29	50.28	206.68	43.27	37.02
207.65	58.48	48.22	205.68	41.43	35.95
206.70	57.15	47.18	204.68	38.73	33.92
205.76	55.39	45.79	203.67	37.21	32.34
204.81	53.98	44.37	202.66	35.14	30.70
203.86	52.48	42.92	201.63	32.93	29.00
202.91	51.18	41.57	200.59	30.55	26.21
201.94	49.37	40.28	199.53	27.95	24.35
200.97	47.85	38.22	198.44	24.89	22.18
199.98	46.06	37.31	197.32	22.56	20.18
198.97	44.46	35.97	196.17	19.35	17.83
197.94	42.38	34.39	194.97	15.01	15.01
196.88	40.86	32.63	193.72	11.27	11.27
195.78	38.24	30.71	192.40	6.67	6.67
194.65	35.49	28.88	191.00	3.32	3.32
193.48	33.01	26.31	189.50	1.00	1.00
192.24	29.64	24.42	187.87	1.00	1.00
190.95	26.13	21.02	186.09	1.00	1.00
189.57	21.96	18.27	184.10	1.00	1.00
188.09	18.47	14.82	181.82	1.00	1.00
186.48	12.81	9.81	179.12	1.00	1.00

RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
184.72	6.67	6.38	175.76	1.00	1.00
182.75	1.00	1.00	171.21	1.00	1.00
180.50	1.00	1.00	163.75	1.00	1.00
177.82	1.00	1.00			
174.50	1.00	1.00			
169.99	1.00	1.00			
162.58	1.00	1.00			

TABLE 17

Item Bank Achievement Estimates in RIT Units and Their NCE Unit Equivalents Based on Fall and Spring Norms--Grades 7 and 8, Mathematics

GRADE 7			GRADE 8		
RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
261.92	99.00	99.00	261.63	99.00	99.00
254.65	99.00	95.00	254.35	99.00	90.17
250.26	99.00	90.40	249.94	92.95	84.38
247.04	93.70	84.82	246.71	88.79	80.22
244.47	93.33	81.23	244.12	83.85	76.03
242.29	89.67	78.41	241.92	79.78	72.86
240.39	86.80	77.15	240.01	76.48	69.76
238.70	83.19	75.01	238.30	74.39	67.93
237.16	81.64	73.05	236.74	71.59	65.92
235.73	79.57	71.38	235.29	69.99	64.13
234.40	77.91	69.66	233.94	67.81	62.34
233.14	75.76	67.68	232.66	66.48	60.78
231.95	73.92	66.45	231.45	64.88	59.29
230.81	72.29	64.89	230.28	62.73	57.85
229.71	70.77	63.42	229.16	61.11	56.47
228.65	69.43	61.84	228.07	59.65	54.97
227.62	67.74	60.46	227.01	57.92	53.33
226.61	66.12	59.15	225.97	56.41	51.56
225.62	64.38	57.71	224.96	54.59	50.32
224.65	62.67	56.08	223.95	52.79	48.77
223.69	61.16	54.84	222.96	51.06	47.38
222.75	59.85	53.45	221.98	49.34	45.60
221.81	58.65	52.08	221.00	47.62	43.97
220.87	56.88	50.94	220.02	45.86	42.57
219.93	55.21	49.39	219.04	44.10	40.44
219.00	53.98	48.22	218.06	42.30	38.56
218.05	52.40	47.13	217.06	40.44	37.43
217.11	51.18	45.47	216.06	37.90	35.29
216.15	49.56	44.36	215.04	36.45	32.99
215.17	47.57	42.66	214.00	33.51	30.97
214.19	45.92	40.91	212.94	31.03	28.58
213.18	43.80	38.73	211.84	28.31	26.37
212.14	41.79	37.65	210.72	26.13	23.52
211.08	38.71	35.24	209.55	22.38	21.13
209.98	36.88	33.06	208.33	18.96	17.34
208.83	33.70	30.68	207.05	14.90	13.96
207.64	30.59	27.67	205.69	12.23	12.23
206.38	27.02	24.92	204.23	6.67	6.67
205.04	23.68	21.18	202.66	2.47	2.47
203.61	18.37	16.60	200.94	1.00	1.00
202.06	13.31	13.26	199.02	1.00	1.00
200.36	8.21	8.21	196.83	1.00	1.00
198.46	1.00	1.00	194.23	1.00	1.00
196.29	1.00	1.00	190.99	1.00	1.00

RIT	FALL NCE	SPRING NCE	RIT	FALL NCE	SPRING NCE
193.71	1.00	1.00	186.57	1.00	1.00
190.49	1.00	1.00	179.29	1.00	1.00
186.09	1.00	1.00			
178.83	1.00	1.00			

students in grades four through eight were given reading and mathematics tests developed from these banks. The ESEA Title I evaluation report for 1978-79 submitted by the Portland School District included average fall and spring achievement estimates and gains in RIT units. This information can be seen in Tables 18 and 19 for Title I and non-Title I students, respectively. In order to convert the fall and spring RIT scores into NCE units, a link to national norms must be established. Portland determined these equivalents from a joint administration of the Comprehensive Tests of Basic Skills, Form S (Comprehensive Tests of Basic Skills, 1975) and selected subsets of bank items. The NCE equivalents reported can also be seen in Tables 18 and 19 (Holmes and Van Cleave, 1979).

A second link to national norms is provided by Tables 12 through 17. In order to use these tables, however, the date of the test administration must conform to the empirical norm dates of the CAT/C, the midpoints being given as October 3 and April 4 (California Achievement Tests, 1979). The Portland fall data were collected during the first three weeks of October, a time interval reasonably close to the fall midpoint date. The spring data, however, were collected during the first three weeks of May, at least one month later than recommended. In order to compensate for this difference, the spring averages were adjusted by determining the average RIT growth from fall to spring at each grade, and then subtracting one seventh of this value from the average spring RIT score. NCE equivalents for the fall and adjusted spring average RIT scores were determined, through interpolation, from Tables 12 through 17. The fall and adjusted spring RIT scores, as well as their NCE equivalents and gains, can be found in Tables 20 and 21 for Title I and non-Title I students, respectively.

TABLE 18

Average Title I Fall 1978 and Spring 1979 Item Bank RIT Achievement Estimates and Their NCE Equivalents as Reported by Portland School District 1J

READING

Grade	N	Fall RIT	Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	720	177.85	187.52	9.67	39.0	44.1	5.1
5	634	185.57	191.77	6.20	38.3	43.0	4.7
6	513	190.26	196.45	6.19	37.1	42.5	5.4
7	453	195.74	199.74	4.00	37.1	40.1	3.0
8	456	198.35	202.48	4.13	35.1	37.7	2.6

MATHEMATICS

Grade	N	Fall RIT	Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	724	180.05	189.70	9.65	37.1	42.5	5.4
5	636	188.16	196.27	8.11	21.8	35.1	13.3
6	519	194.31	201.62	7.31	28.2	38.3	10.1
7	456	200.20	205.13	4.93	13.1	26.3	13.2
8	458	204.90	209.61	4.71	24.2	34.4	10.2

TABLE 19

Average Non-Title I Fall 1978 and Spring 1979 Item Bank RIT Achievement Estimates and Their NCE Equivalents as Reported by Portland School District 1J

READING

Grade	N	Fall RIT	Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	2,660	197.78	204.03	6.25	64.2	67.0	2.8
5	2,564	205.02	209.73	4.71	66.3	67.0	.7
6	2,692	209.74	214.35	4.61	63.5	65.6	2.1
7	2,625	215.09	219.79	4.70	64.9	68.5	3.6
8	2,759	2[illegible]12	224.85	4.73	65.6	70.9	5.3

MATHEMATICS

Grade	N	Fall RIT	Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	2,659	194.43	204.64	10.21	53.2	62.3	9.1
5	2,561	204.85	212.36	7.51	57.0	59.3	2.3
6	2,701	211.73	217.50	5.77	54.8	60.4	5.6
7	2,645	219.13	224.20	5.07	58.7	59.9	1.2
8	2,795	225.08	228.71	3.63	57.5	58.1	.6

TABLE 20

Average Title I Fall 1978 and Adjusted Spring 1979 Item Bank RIT Achievement Estimates for Portland School District 1J and Their NCE Equivalents as Determined from Project Conversion Tables

READING

Grade	N	Fall RIT	Adjusted Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	720	177.85	186.14	8.29	23.16	35.81	12.65
5	634	185.57	190.88	5.31	28.01	33.65	5.64
6	513	190.26	195.57	5.31	24.92	32.08	7.16
7	453	195.74	199.17	3.43	17.65	26.41	8.76
8	456	198.35	201.89	3.54	6.23	16.85	10.62

MATHEMATICS

Grade	N	Fall RIT	Adjusted Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	724	180.05	188.32	8.27	18.60	30.55	11.95
5	636	188.16	195.11	6.95	18.64	29.62	10.98
6	519	194.31	200.58	6.27	13.04	26.69	13.65
7	456	200.20	204.43	4.23	6.60	19.23	12.63
8	458	204.90	208.94	4.04	9.22	19.24	10.01

TABLE 21

Average Non-Title I Fall 1978 and Adjusted Spring 1979 Item Bank RIT Achievement Estimates for Portland School District 1J and Their NCE Equivalents as Determined from Project Conversion Tables

READING

Grade	N	Fall RIT	Adjusted Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	2,660	197.78	203.14	5.36	56.93	55.02	-1.91
5	2,564	205.02	209.06	4.04	58.08	56.85	-1.23
6	2,692	209.74	213.69	3.95	56.20	56.43	.23
7	2,625	215.09	219.12	4.03	57.79	59.02	1.23
8	2,759	220.12	224.17	4.05	57.16	59.04	1.88

MATHEMATICS

Grade	N	Fall RIT	Adjusted Spring RIT	RIT Gain	Fall NCE	Spring NCE	NCE Gain
4	2,659	194.43	203.18	8.75	50.70	53.34	2.64
5	2,561	204.85	211.29	6.44	54.04	54.32	.28
6	2,701	211.73	216.68	4.95	53.11	53.40	.29
7	2,645	219.13	233.48	4.35	54.15	54.53	.38
8	2,795	225.08	228.19	3.11	54.81	55.14	.33

The average RIT score at each grade was computed for the Project sample. These values can be seen in Table 22.

TABLE 22

Average Achievement Estimates in RIT Units for Project Sample by Content Area and Grade

GRADE	N[a]	READING	N[a]	MATHEMATICS
3	456	195.2	341	193.1
4	457	203.6	354	203.1
5	479	209.7	370	210.3
6	460	212.7	428	218.2
7	479	219.1	546	225.7
8	559	222.5	553	230.5

[a]Differences in the sample sizes reported here and those appearing in Table 3 are due to the students with all items correct or incorrect.

RIT and Expanded Scale Score Comparison

In educational measurement, it is frequently necessary to make statements about the performance of students who took different levels of the same test. In order to facilitate these kinds of comparisons, many test publishers attempt to construct a score scale which spans all levels of their test. Such an expanded score scale has been developed for the CAT/C using Thurstone's absolute scaling procedure (California Achievement Tests, 1979). This single, equal-interval scale ranges in value from 1 to 999, with the scale's mean and standard deviation being anchored at 600 and 90, respectively,

for the grade 10 group. Since equivalent CAT/C converted raw scores have been determined through Rasch model equating for each item bank raw score, the relationship between the RIT and expanded score scales can easily be developed by expressing these raw score values in terms of their equal-interval scale values. Figures 1 and 2 present these relationships for reading and mathematics, respectively.

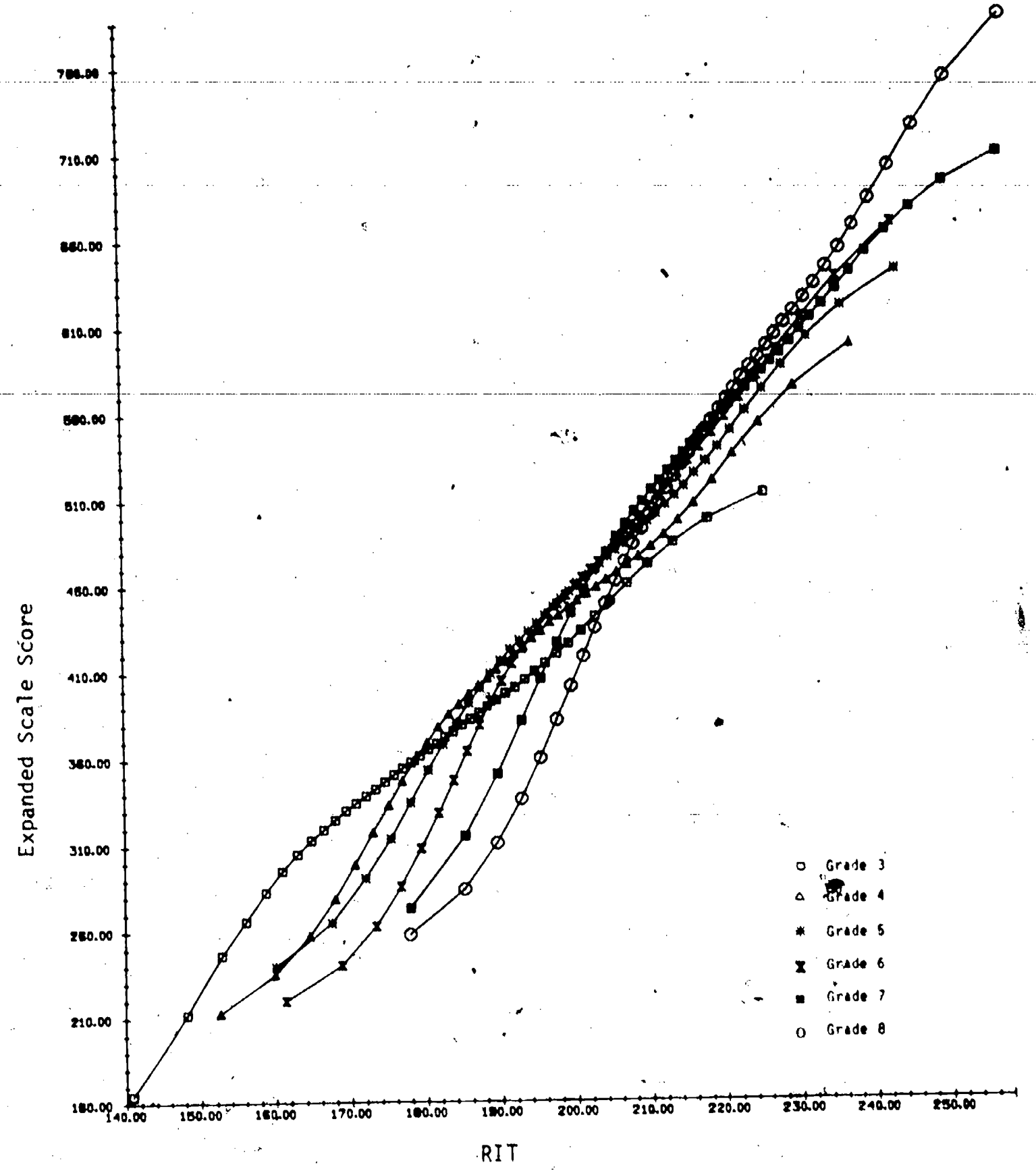

Figure 1

Graph of the Relationship Between RIT and Expanded Scale Scores in Reading

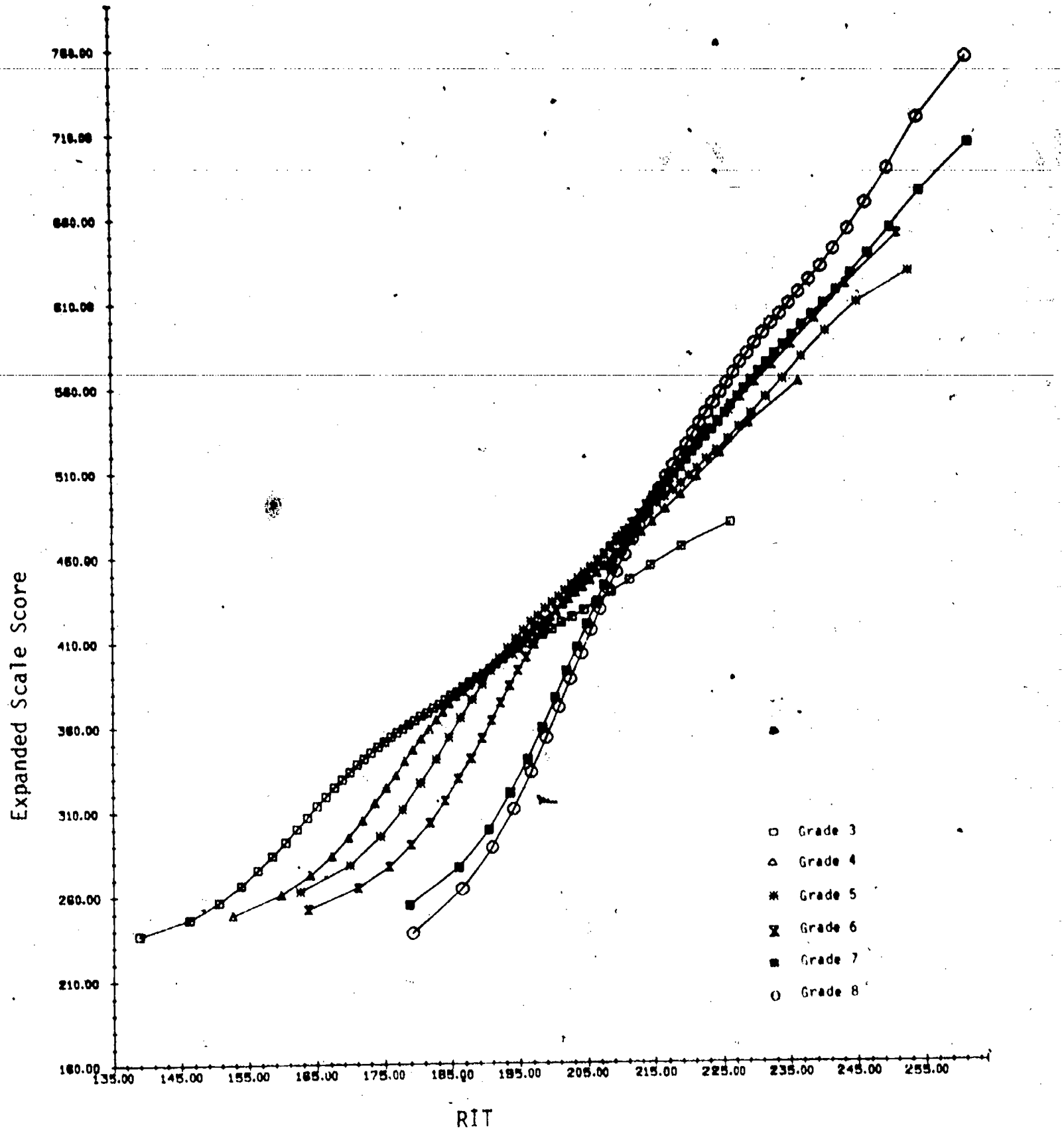

Figure 2

Graph of the Relationship Between RIT and Expanded Scale Scores in Mathematics

DISCUSSION

Methodological Considerations

Before any discussion of the Project findings can be meaningfully conducted, a few general comments related to methodology seem in order. First, it should be noted that the test equating design adopted by the Project, that of administering both tests to the same group of students, is a fairly strong one. One possible improvement might have been to counterbalance the order of the two test administrations, thus eliminating possible fatigue or practice effects. This was not feasible, since the data were collected as part of routine district-wide evaluation programs. The design used, however, still represents a highly desirable testing procedure for an equating study.

Second, the Project sample, while not necessarily identical to the CAT/C norming group, did demonstrate a consistent relationship to it across all grades and both content areas studied. A comparison of the CAT/C raw score means and standard deviations presented in Table 6 with those reported for the CAT/C spring norming sample (California Achievement Tests, 1979) indicated that the performance of the Project sample was slightly higher and more homogeneous than that of the norming sample. These differences were, however, consistent across all grades and not unexpected given the purpose of the CAT/C sampling strategy. In addition, the average achievement estimates reported in Table 22 for the Project sample were consistent with those reported in Tables 20 and 21 for the spring testing of Portland Public School students.

A third area requiring comment is that of instrumentation. Since only total test equating was considered in the Project, the correlations between

total test item bank and CAT/C raw scores are of interest. One of the assumptions inherent in test equating is that both tests measure the same trait. While the total test correlations reported in Table 7 are lower than might be expected, the larger corrected values also reported there suggest that an adequate relationship between the equating test pair members did exist. An interesting observation related to the subtest raw score correlations presented in Table 6 of Appendix A should be noted. The expected increase in homogeneity between subtest raw scores is not reflected in the subtest correlations reported there; that is, item bank and CAT/C total test correlations are consistently higher than item bank and CAT/C subtest correlations. This suggests a stronger relationship between total test scores than between subtest scores, thus lending support to the total test equating approach utilized.

Some rather interesting results related to model data fit were reported in Table 8. While no absolute criterion exists for determining when the assumption of equal item discrimination is no longer tenable, the percents of nonfitting items reported are disturbingly high. It has been demonstrated that when this assumption is violated, discrepancies in item difficulty estimates may result, and that these discrepancies can adversely effect achievement estimates obtained from vertical equating situations (George, 1979). The results reported in Table 8 suggest that many items on both the item bank and CAT/C reading and mathematics tests do not satisfy this assumption, especially at the upper grade levels. Since the robustness of the model to these violations is still an empirical question, the implications of these findings are unclear. It should, however, be noted that equal item discrimination is only one aspect of model data fit. Further research is needed in this area.

Equating Comparisons

An examination of the results of the four equating techniques leads to the conclusion that, for this sample, all four procedures appear to work equally well. This becomes most apparent in Figures 1 through 12 in Appendix C. It should be noted that no smoothing techniques were applied to the equipercentile results, thus accounting for some of the abnormalities seen in these curves. This is especially true at the lower end of the distributions where, frequently, no actual data points were observed. The overall average absolute discrepancy between converted and observed CAT/C scores reported in Table 9 is a further indication of the similarity of the four sets of equating results. While the repeated measures analyses of variance suggest significant differences between these averages at certain grades, the analyses of effects size indicate that these differences are more a function of the large number of observations than any real technique differences. In all cases, the amount of variance accounted for was less than one percent.

One disturbing aspect of the overall average absolute discrepancy values reported in Table 9 is their size. These values are reported in CAT/C raw score points, suggesting that any converted CAT/C score may, on the average, differ by from five to seven CAT/C raw score points from the value actual observed in the Project sample. This discrepancy is alarmingly large since the standard errors of measurement for the CAT/C tests are roughly 3.5 points. Further, an examination of these discrepancies by item bank raw scores (Appendix D) indicates that the largest discrepancy values occur in the lower end of the raw score distributions. This appears to be the case regardless of the equating technique used, suggesting that converted CAT/C scores at the lower end of the score

distribution are consistently less accurately determined. The large discrepancies observed may be a result of only moderately high correlations between equating test pair members. This fact may be further compounded by the general instability of test scores in the lower extremes of the score distribution. Regardless of the reason for the large discrepancies, the lower end of the distribution is unfortunately where most Title I students may be expected to score.

In order to determine the nature of the Rasch model equating discrepancies reported in Table 9, plots of their distributions were included in Appendix E. The correlations between these values and their corresponding item bank raw scores appear in Table 11. Two general conclusions can be drawn from these data. First, the plots confirm the fact that the largest discrepancies occur at the lower end of the raw score distributions. They further suggest that scores in the lower end are generally underestimated while scores at the upper end tend to be overestimated by this equating technique. This is further supported by the negative correlations in Table 11. While the size of the correlations is small, all values, except grade 7 reading, were statistically significant. These data further suggest a need for caution in interpreting scores in the lower end of the raw score distribution.

RIT to NCE Conversions

The results obtained by applying the Project developed conversion tables to the data provided by the Portland School District can be found in Tables 20 and 21 for Title I and non-Title I students, respectively. The Title I fall and spring equivalents are consistent with the definition of Title I students as lower achievers. It is interesting to note

the pattern of decreasing fall NCE equivalents from the lower to upper grades in both reading and mathematics. While the NCE gains from fall to spring are in many cases quite large, Title I students appear to lose ground rapidly as they proceed through the grades, beginning each year at a successively lower NCE level than the year before. This pattern may, however, be more indicative of a basic change in the Title I population than an actual decrease in achievement. It probably reflects the movement of remediated students into non-Title I programs, leaving behind a smaller "hard core" group of Title I students each year.

The values of the non-Title I fall and spring equivalents are somewhat higher than might be expected, especially in reading. This inflation is due in part to the fact that the performance of Title I students was not included in the computation of these RIT achievement estimates. It appears that while Title I students begin each school year at successively lower NCE levels, non-Title I students appear to begin at successively higher levels.

The pattern of NCE gains for non-Title I students is rather interesting. Unlike the RIT gains for this group, reading NCE gains are initially negative, becoming increasingly more positive in the upper grades. With the exception of the grade 4 value, a similar tendency is suggested by the mathematics data. This pattern is the opposite of what might be expected, given the familiar learning curve where the largest gains occur early in the learning experience. In addition, since an NCE gain of 0.00 represents normal growth from fall to spring, both positive and negative gains would be expected to occur. This particular pattern of gains suggests increasingly more effective reading and mathematics instruction at the upper grades.

A second set of NCE equivalents, developed by the Portland School District through the California Tests of Basic Skills, are reported in Tables 18 and 19 for Title I and non-Title I students, respectively. In general, the fall and spring NCE equivalents determined by Portland are consistently higher in both content areas than those developed by this Project. While the two sets of non-Title I gains are somewhat similar, the Title I NCE gains differ considerably. The only consistent relationship displayed by the two sets of Title I NCE values is the decline in both fall and spring equivalents across grades noted earlier.

The lack of similarity between the two sets of NCE equivalents is disturbing, especially since both were developed using similar techniques. The generally higher NCE values reported by the Portland School District may reflect basic differences between the norms of the two instruments used in the determination of the equivalents. This is a common problem encountered when making norm-referenced score interpretations; that is, the performance of the same student, or group of students, may appear to be quite different when evaluated by two different norm-referenced instruments which claim to measure the same thing. Of particular concern, however, are the large discrepancies between the two sets of Title I NCE gains. This further lends support to the earlier assertion that caution must be exercised when attempting norm-referenced interpretation of scores found in the lower end of the score distribution.

The RIT and Expanded Score Scale

A few comments concerning the relationship between the RIT and expanded score scale are in order. No attempt was made to empirically evaluate the relationship. While extremely interesting, this was beyond the scope

of the Project. However, the plots presented in Figures 1 and 2 do suggest that, at least within the mid-range of each grade, the relationship across grades is somewhat linear. It should be noted that the slope of the lines representing the relationship at each grade changes rather consistently across grades. This phenomena can be seen in both Figures 1 and 2. This suggests that a change in the unit of measure of one or both of the scales is occurring, making their equal-interval claim questionable.

CONCLUSIONS

The purpose of the Project was to explore an alternative approach to the implementation of Model A for Title I evaluation, namely the use of Rasch calibrated item banks. The methodology described in this report along with the conversion tables for interpreting item bank achievement estimates in terms of NCE units provide support for the feasibility of such an alternative.

Three conventional equating techniques in addition to a Rasch model approach were considered during the development of the Project conversion tables. While no absolute criterion exists for evaluating the accuracy of equating results, all three techniques compared favorably to the Rasch model approach when comparisons were based on the performance of the Project sample.

Although all four equating techniques yielded similar results, the accuracy of the results produced, especially for scores in the lower end of the score distribution, was a concern. The average absolute discrepancy between scores actually observed and those determined from the equating process was rather large. The largest values occurred in the lower end of the distribution, suggesting that the least accurate results will be obtained by applying the Project conversion tables to the data from those students for which the tables were developed.

This situation was further reflected in the disparity between Title I NCE gains reported by Portland Public Schools and those determined by this Project for the same data. Both sets of gains were developed from a similar test equating approach but using different norm-referenced tests.

It was suggested that this inconsistency might be due to differences between the norms for the two tests. However, it was concluded that, while the implementation strategy for Model A studied here was feasible, the accuracy of the results it produced was questionable.

REFERENCES

Andersen, E. B. A goodness of fit test for the Rasch model. Psychometrika, 1973, 38, 123-140.

Angoff, W. H. Scales, norms, and equivalent scores. In Thorndike (ed.), Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

California Achievement Tests: Norms Tables. Monterey, CA: CTB/McGraw-Hill, 1978.

California Achievement Tests: Technical Bulletin 1. Monterey, CA: CTB/McGraw-Hill, 1979.

Comprehensive Tests of Basic Skills. Monterey, CA: CTB/McGraw-Hill, 1975.

George, A. A. Theoretical and practical consequences of the use of standardized residuals as Rasch model fit statistics. A paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Gustafsson, J. E. Testing and obtaining fit of data to the Rasch model. A paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Forster, F. & Doherty, V. Using the Rasch model to improve your district's testing program. A paper presented at the meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Hambleton, R. K. & Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eigner, D. R. & Gifford, J.A. Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 1978, 48, 467-510.

Holmes, J. N. & Van Cleave, M. E. Area I Disadvantaged Child Project, Evaluation report, ESEA, Title I, 1978-79. Portland, OR: Portland School District 1J, August 1979.

Lord, F. M. Notes on comparable scales for test scores (ETS RB 50-48). Princeton, N.J.: Educational Testing Service, 1950.

Lord, F. M. A theory of test scores. Psychometric Monograph (No. 7), 1952.

Lord, F. M. The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 1953, 13, 517-548.

Lord, F. M. Evaluation with artificial data of a procedure for estimating ability and ten characteristic curve parameters (ETS RB 75-33). Princeton, N.J.: Educational Testing Service, 1975.

Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 177-138.

Lord, F. M. & Novick, M. R. Statistical theories of mental curve theory. Reading, MA: Addison-Wesley, 1968.

Northwest Evaluation Association. Rasch model monograph series. Portland, OR: Author, December 1978.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedogogiske Institute, 1960.

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57.

Rentz, R. R. & Bashaw, W. L. Equating reading tests with the Rasch model, Volume I final report, Volume II technical reference tables. Athens, GA: University of Georgia, Educational Research Laboratory, 1975.

Rentz, R. R. & Bashaw, W. L. The national reference scale for reading: An application of the Rasch model. Journal of Educational Measurement, 1977, 14, 161-180.

Tallmadge, G. K. & Wood, C. T. User's guide, ESEA Title I, Evaluation and reporting system. Mountain View, CA: RMC Research Corporation, January 1978.

Wichert, V. E. Methods of equating test forms and an equating computer system. Princeton, NJ: Educational Testing Service, December 1976.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Wright, B. D. Sample-free test calibration and person measurement. Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1968, 85-101.

Wright, B. D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Wright, B. D. & Douglas, G. A. Best procedures for sample-free item analysis. Applied Psychological Measurement, 1977, 1, 281-295.

Wright, B. D. & Mead, R. J. BICAL: Calibrating items and scales with the Rasch model (Research Memorandum No. 23). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago, 1977.

Wright, B. D., Mead, R. & Draba, R. Detecting and correcting item bias with a logical response model (Research Memorandum No. 22). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago, 1976.

Wright, B. D. & Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Wright, B. D. & Stone, M. H. Best test design: A handbook for Rasch measurement. Chicago, IL: Mesa Press, 1979.

APPENDIX A

Subtest Summary Statistics

TABLE 1

Number and Percent of Items Measuring Reading Subskills on Each Equating Test Pair After Item Deletion by Grade

Grade	Test	Vocabulary	Comprehension	Total Remaining
3	CAT13[a]	15 (21%)	27 (37%)	73
	Item Bank Test (318)	15 (34%)	29 (66%)	44
4	CAT14	30 (43%)	40 (57%)	70
	Item Bank Test (319)	11 (26%)	31 (74%)	42
5	CAT15	30 (43%)	40 (57%)	70
	Item Bank Test (320)	12 (27%)	33 (73%)	45
6	CAT16	30 (43%)	40 (57%)	70
	Item Bank Test (320)	11 (26%)	31 (74%)	42
7	CAT17	30 (43%)	40 (57%)	70
	Item Bank Test (321)	10 (23%)	33 (77%)	43
8	CAT18	30 (43%)	40 (57%)	70
	Item Bank Test (321)	10 (23%)	34 (77%)	44

[a] CAT13 also contains a twenty item Phonic Analysis subtest and an eleven item Structural Analysis subtest, representing 27 percent and 15 percent of the total number of items respectively.

TABLE 2

Number and Percent of Items Measuring Mathematics Subskills on Each Equating Test Pair After Item Deletion by Grade

Grade	Test	Computations	Concepts	Total Remaining
3	CAT13	40 (47%)	45 (53%)	85
	Item Bank Test (818)	20 (37%)	34 (63%)	54
4	CAT14	40 (47%)	45 (53%)	85
	Item Bank Test (819)	21 (40%)	32 (60%)	53
5	CAT15	40 (47%)	45 (53%)	85
	Item Bank Test (820)	22 (42%)	30 (58%)	52
6	CAT16	40 (47%)	45 (53%)	85
	Item Bank Test (820)	21 (44%)	27 (56%)	48
7	CAT17	40 (47%)	45 (53%)	85
	Item Bank Test (821)	20 (41%)	29 (59%)	49
8	CAT18	40 (47%)	45 (53%)	85
	Item Bank Test (821)	20 (43%)	27 (57%)	47

TABLE 3

Reading Vocabulary and Comprehension Subtest Raw Score Means, Standard Deviations and Reliability Estimates for the Item Bank and CAT/C Tests by Grade

	ITEM BANK						CAT/C					
		Vocabulary			Comprehension			Vocabulary			Comprehension	
GRADE	N	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20
3[a]	463	11.1	3.6	.83	19.9	5.8	.87	12.4	2.8	.81	22.3	4.5	.86
4	457	7.0	2.3	.65	20.4	5.9	.84	23.3	6.1	.89	26.2	8.2	.90
5	481	7.9	2.9	.74	20.9	6.2	.85	24.1	4.8	.85	27.6	8.2	.91
6	462	7.9	2.5	.72	20.7	6.1	.86	21.3	5.5	.86	27.0	7.7	.89
7	479	5.5	2.2	.58	17.5	5.8	.80	21.5	5.2	.83	25.3	8.0	.89
8	559	6.2	2.2	.62	20.0	6.4	.83	18.1	6.8	.89	25.7	7.0	.86

[a]The raw score mean, standard deviation and reliability estimate for the third grade Phonetics Analysis and Structural Analysis subtests were 15.9, 3.8, .84, and 8.6, 2.0 and .67 respectively.

TABLE 4

Mathematics Computations and Concepts Subtest Raw Score Means, Standard Deviations and Reliability Estimates for the Item Bank and CAT/C Tests by Grade

	ITEM BANK						CAT/C					
	Computations			Concepts			Computations			Concepts		
GRADE	N	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20	MEAN	STANDARD DEVIATION	KR-20
3	343	13.1	4.6	.85	23.8	6.2	.86	25.3	8.0	.91	31.2	8.0	.89
4	354	14.1	5.0	.86	20.7	6.6	.87	24.0	8.3	.90	28.0	8.8	.90
5	370	11.4	4.7	.83	17.6	6.0	.84	24.7	7.7	.88	27.3	9.1	.90
6	429	14.4	4.7	.84	18.6	4.9	.83	24.9	7.4	.88	27.3	8.2	.88
7	549	12.1	4.6	.83	17.1	6.2	.86	22.1	8.8	.91	26.5	8.7	.90
8	559	13.5	4.5	.83	18.1	5.8	.86	26.4	8.8	.92	28.6	9.0	.90

TABLE 5

Item Bank and CAT/C Within Test Subtest Correlations Uncorrected and Corrected for Attenuation by Content Area and Grade

READING

	ITEM BANK		CAT/C	
GRADE	r[a]	r corrected for attenuation	r[a]	r corrected for attenuation
3	.84	.99	.72	.87
4	.72	.97	.79	.89
5	.75	.95	.70	.80
6	.76	.97	.78	.89
7	.67	.98	.75	.87
8	.72	1.00	.79	.90

MATHEMATICS

GRADE	r[b]	r corrected for attenuation	r[b]	r corrected for attenuation
3	.74	.90	.66	.73
4	.80	.92	.75	.83
5	.73	.87	.71	.80
6	.73	.88	.77	.87
7	.77	.91	.78	.86
8	.74	.87	.80	.88

[a]Vocabulary vs. comprehension

[b]Concepts vs. computations

TABLE 6

Item Bank and CAT/C Between Test Subtest Correlations Uncorrected and Corrected by Content Area and Grade

READING

	VOCABULARY		COMPREHENSION	
GRADE	r	r corrected for attenuation	r	r corrected for attenuation
3	.59	.72	.63	.73
4	.67	.88	.77	.89
5	.66	.84	.78	.89
6	.72	.91	.81	.92
7	.63	.90	.73	.87
8	.66	.89	.80	.94

MATHEMATICS

	COMPUTATIONS		CONCEPTS	
GRADE	r	r corrected for attenuation	r	r corrected for attenuation
3	.82	.93	.82	.93
4	.80	.90	.82	.93
5	.75	.88	.80	.92
6	.81	.95	.79	.93
7	.79	.91	.84	.96
8	.79	.90	.77	.87

APPENDIX B

CAT/C Converted Scores Yielded by Four Equating Methods
for Scores on the Corresponding Item Bank Test

TABLE 1

CAT13 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (B18) - Grade 3, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT13 Converted Score LO	LT	EQ	RM
43	7	73.86	73.97	72.54	72.17
42	16	72.63	72.73	71.68	71.30
41	24	71.41	71.49	70.79	70.40
40	15	70.18	70.26	69.84	69.47
39	29	68.95	69.02	68.69	68.51
38	29	67.72	67.78	67.55	67.52
37	26	66.49	66.54	66.51	66.50
36	26	65.26	65.30	65.38	65.46
35	27	64.03	64.07	63.96	64.40
34	30	62.80	62.83	62.21	63.31
33	16	61.57	61.59	60.98	62.20
32	27	60.35	60.35	59.93	61.07
31	7	59.12	59.12	59.11	59.92
30	20	57.89	57.88	58.32	58.75
29	17	56.66	56.64	57.06	57.56
28	12	55.43	55.40	56.06	56.35
27	11	54.20	54.17	55.19	55.13
26	12	52.97	52.93	54.05	53.88
25	6	51.74	51.69	52.76	52.62
24	3	50.51	50.45	52.01	51.34
23	9	49.29	49.22	51.16	50.03
22	6	48.06	47.98	50.15	48.70
21	5	46.83	46.74	49.33	47.35
20	9	45.60	45.50	48.12	45.98
19	5	44.37	44.27	46.86	44.58
18	3	43.14	43.03	46.13	43.11
17	5	41.91	41.79	45.07	41.68
16	9	40.68	40.55	42.37	40.18
15	8	39.45	39.31	40.42	38.65
14	8	38.23	38.08	38.67	37.07
13	8	37.00	36.84	35.20	35.44
12	8	35.77	35.60	32.77	33.75
11	4	34.54	34.36	29.70	32.01
10	2	33.31	33.13	28.82	30.20
9	2	32.08	31.89	28.16	28.32
8	3	30.85	30.65	26.99	26.33
7	1	29.62	29.41	20.49	24.24
6	1	28.40	28.18	19.50	22.01
5	0	27.17	26.94	19.00	19.61
4	0	25.94	25.70	19.00	17.00
3	0	24.71	24.46	19.00	14.07
2	0	23.48	23.23	19.00	10.69
1	0	22.25	21.99	19.00	6.47

TABLE 2

CAT14 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (319) - Grade 4, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT14 Converted Score LO	LT	EQ	RM
41	3	73.54	74.41	69.25	68.95
40	11	71.77	72.58	67.79	67.83
39	12	70.01	70.75	66.75	66.65
38	14	68.24	68.91	65.94	65.41
37	15	66.47	67.08	65.07	64.13
36	13	64.70	65.25	64.20	62.80
35	17	62.93	63.42	63.15	61.43
34	28	61.16	61.58	61.68	60.03
33	26	59.40	59.75	60.02	58.59
32	19	57.63	57.92	58.07	57.13
31	23	55.86	56.09	56.66	55.63
30	20	54.09	54.25	54.86	54.12
29	20	52.32	52.42	53.35	52.58
28	22	50.55	50.59	52.03	51.02
27	21	48.78	48.76	50.27	49.44
26	27	47.02	46.92	48.05	47.84
25	12	45.25	45.09	46.19	46.22
24	20	43.48	43.26	44.41	44.59
23	16	41.71	41.43	42.03	42.94
22	11	39.94	39.59	40.45	41.28
21	14	38.17	37.76	38.96	39.60
20	13	36.41	35.93	36.59	37.90
19	1	34.64	34.10	34.38	36.19
18	13	32.87	32.27	31.98	34.47
17	11	31.10	30.43	29.63	32.73
16	7	29.33	28.60	26.80	30.98
15	10	27.56	26.77	25.32	29.21
14	6	25.80	24.94	24.07	27.43
13	6	24.03	23.10	22.84	25.64
12	8	22.26	21.27	20.71	23.82
11	3	20.49	19.44	18.79	22.00
10	2	18.72	17.61	17.83	20.15
9	2	16.95	15.77	17.16	18.29
8	1	15.18	13.94	16.35	16.41
7	1	13.42	12.11	12.99	14.52
6	0	11.65	10.28	12.00	12.60
5	0	9.88	8.44	12.00	10.65
4	0	8.11	6.61	12.00	8.68
3	0	6.34	4.78	12.00	6.68
2	0	4.57	2.95	12.00	4.62
1	0	2.81	1.11	12.00	2.46

TABLE 3

CAT15 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (320) - Grade 5, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT15 Converted Score LO	LT	EQ	RM
44	4	73.18	73.66	69.56	69.05
43	5	71.77	72.22	68.77	68.09
42	12	70.36	70.78	67.77	67.11
41	10	68.95	69.34	66.69	66.11
40	21	67.54	67.90	65.68	65.10
39	20	66.13	66.45	64.45	64.08
38	18	64.72	65.01	63.39	63.03
37	13	63.31	63.57	62.62	61.98
36	24	61.90	62.13	61.68	60.90
35	16	60.49	60.69	60.50	59.81
34	13	59.08	59.25	59.65	58.70
33	22	57.68	57.81	58.71	57.57
32	17	56.27	56.37	57.47	56.43
31	23	54.86	54.93	55.89	55.26
30	20	53.45	53.49	54.47	54.07
29	20	52.04	52.05	53.36	52.87
28	21	50.63	50.61	52.19	51.63
27	18	49.22	49.18	50.31	50.38
26	14	47.81	47.72	49.08	49.10
25	14	46.40	46.28	47.90	47.80
24	18	44.99	44.84	46.64	46.47
23	17	43.58	43.40	44.76	45.11
22	13	42.17	41.96	43.09	43.73
21	6	40.76	40.52	42.17	42.31
20	13	39.36	39.08	41.38	40.87
19	15	37.95	37.64	40.04	39.39
18	12	36.54	36.20	37.52	37.87
17	10	35.13	34.76	35.09	36.32
16	12	33.72	33.31	32.90	34.73
15	9	32.31	31.87	30.29	33.10
14	7	30.90	30.43	27.27	31.42
13	9	29.49	28.99	25.74	29.70
12	4	28.08	27.55	22.47	27.93
11	3	26.67	26.11	20.59	26.10
10	2	25.26	24.67	19.33	24.21
9	2	23.85	23.23	17.99	22.26
8	1	22.44	21.79	16.99	20.24
7	0	21.04	20.35	16.00	18.15
6	1	19.63	18.91	12.99	15.96
5	0	18.22	17.46	12.00	13.68
4	0	16.81	16.02	12.00	11.29
3	0	15.40	14.58	12.00	8.76
2	0	13.99	13.14	12.00	6.08
1	0	12.58	11.70	12.00	3.19

TABLE 4

CAT16 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (320) - Grade 6, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT16 Converted Score LO	LT	EQ	RM
41	8	67.13	67.49	67.87	68.07
40	15	65.60	65.93	65.90	66.23
39	23	64.07	64.38	64.20	64.47
38	20	62.55	62.82	62.36	62.76
37	22	61.02	61.26	60.78	61.10
36	25	59.49	59.71	58.97	59.48
35	27	57.96	58.15	57.35	57.88
34	21	56.44	56.59	55.85	56.31
33	21	54.91	55.04	54.24	54.76
32	16	53.38	53.48	52.72	53.23
31	29	51.86	51.93	51.17	51.71
30	16	50.33	50.37	49.74	50.20
29	18	48.80	48.81	48.69	48.69
28	14	47.28	47.26	47.56	47.19
27	11	45.75	45.70	46.52	45.70
26	12	44.22	44.15	45.33	44.20
25	12	42.70	42.59	44.10	42.70
24	22	41.17	41.03	42.32	41.20
23	12	39.64	39.48	40.74	39.69
22	14	38.12	37.92	39.33	38.18
21	15	36.59	36.36	37.67	36.67
20	11	35.06	34.81	35.85	35.14
19	8	33.54	33.25	34.53	33.61
18	11	32.01	31.70	33.33	32.06
17	9	30.48	30.14	31.43	30.51
16	10	28.96	28.58	29.00	28.94
15	9	27.43	27.03	27.00	27.36
14	12	25.90	25.47	24.40	25.76
13	2	24.38	23.92	21.62	24.14
12	3	22.85	22.36	21.00	22.51
11	4	21.32	20.80	19.60	20.86
10	4	19.80	19.25	17.50	19.18
9	2	18.27	17.69	15.33	17.48
8	2	16.74	16.13	13.00	15.75
7	0	15.21	14.58	12.00	13.99
6	0	13.69	13.02	12.00	12.20
5	0	12.16	11.47	12.00	10.36
4	0	10.63	9.91	12.00	8.47
3	0	9.11	8.35	12.00	6.52
2	0	7.58	6.80	12.00	4.48
1	0	6.05	5.24	12.00	2.32

TABLE 5

CAT17 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (321) - Grade 7, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT17 Converted Score LO	LT	EQ	RM
42	1	78.14	79.60	69.75	69.11
41	2	76.49	77.88	69.00	68.19
40	0	74.85	76.16	68.61	67.25
39	1	73.21	74.44	68.41	66.29
38	7	71.56	72.72	67.16	65.31
37	2	69.92	71.00	66.59	64.30
36	7	68.28	69.28	66.12	63.28
35	10	66.63	67.56	65.44	62.23
34	15	64.99	65.84	64.25	61.16
33	12	63.35	64.12	62.38	60.06
32	17	61.70	62.40	60.74	58.94
31	11	60.06	60.68	59.38	57.81
30	16	58.42	58.96	58.33	56.65
29	23	56.77	57.24	56.65	55.46
28	20	55.13	55.52	54.92	54.25
27	21	53.49	53.80	53.55	53.02
26	17	51.84	52.08	52.28	51.76
25	13	50.20	50.36	51.34	50.48
24	27	48.56	48.64	50.08	49.16
23	16	46.91	46.92	48.16	47.82
22	30	45.27	45.20	46.45	46.45
21	29	43.63	43.48	44.19	45.04
20	18	41.98	41.76	42.27	43.61
19	16	40.34	40.04	40.92	42.14
18	25	38.70	38.31	38.97	40.63
17	19	37.05	36.59	37.06	39.08
16	21	35.41	34.87	35.55	37.49
15	9	33.76	33.15	34.33	35.86
14	14	32.12	31.43	33.26	34.18
13	15	30.48	29.71	31.72	32.44
12	14	28.83	27.99	28.36	30.65
11	13	27.19	26.27	25.12	28.79
10	9	25.55	24.55	21.77	26.86
9	3	23.90	22.83	19.48	24.86
8	2	22.26	21.11	15.94	22.76
7	3	20.62	19.39	13.99	20.57
6	1	18.97	17.67	11.99	18.26
5	0	17.33	15.95	11.00	15.81
4	0	15.69	14.23	11.00	13.20
3	0	14.04	12.51	11.00	10.39
2	0	12.40	10.79	11.00	7.32
1	0	10.76	9.07	11.00	3.91

TABLE 6

CAT18 Converted Total Reading Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (321) - Grade 8, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT18 Converted Score LO	LT	EQ	RM
43	0	70.75	71.63	70.00	68.50
42	5	69.14	69.97	67.83	67.01
41	4	67.54	68.31	66.60	65.52
40	8	65.93	66.66	65.60	64.04
39	7	64.33	65.00	64.68	62.57
38	15	62.73	63.34	63.50	61.10
37	22	61.12	61.69	61.42	59.64
36	19	59.52	60.03	59.75	58.18
35	25	57.91	58.37	58.33	56.72
34	23	56.31	56.72	56.50	55.26
33	25	54.71	55.06	54.39	53.81
32	19	53.10	53.41	52.76	52.36
31	25	51.50	51.75	50.96	50.90
30	18	49.89	50.09	49.40	49.45
29	19	48.29	48.44	48.22	48.00
28	25	46.69	46.78	46.61	46.54
27	22	45.08	45.12	45.12	45.09
26	26	43.48	43.47	43.33	43.63
25	29	41.87	41.81	41.38	42.16
24	13	40.27	40.15	39.72	40.70
23	18	38.66	38.50	38.60	39.22
22	15	37.06	36.84	37.40	37.75
21	15	35.46	35.18	36.31	36.26
20	24	33.85	33.53	34.44	34.77
19	18	32.25	31.87	32.59	33.27
18	18	30.64	30.22	31.04	31.76
17	18	29.04	28.56	29.56	30.24
16	19	27.44	26.90	27.67	28.80
15	14	25.83	25.25	25.72	27.16
14	12	24.23	23.59	24.50	25.59
13	19	22.62	21.93	22.15	24.01
12	7	21.02	20.28	19.82	22.42
11	6	19.42	18.62	18.43	20.80
10	2	17.81	16.96	17.17	19.16
9	2	16.21	15.31	16.25	17.49
8	1	14.60	13.65	15.00	15.79
7	0	13.00	12.00	14.50	14.06
6	2	11.40	10.34	13.00	12.29
5	0	9.79	8.68	10.00	10.47
4	0	8.19	7.03	10.00	8.60
3	0	6.58	5.37	10.00	6.65
2	0	4.98	3.71	10.00	4.61
1	0	3.38	2.06	10.00	2.42

CAT13 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (818) - Grade 3, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT13 Converted Score LO	LT	EQ	RM
53	3	79.95	80.26	83.66	82.80
52	4	78.50	78.79	81.66	80.73
51	7	77.04	77.32	78.24	78.78
50	11	75.59	75.84	76.61	76.92
49	8	74.14	74.37	74.49	75.13
48	12	72.69	72.90	73.22	73.40
47	14	71.24	71.43	71.62	71.72
46	13	69.78	69.96	69.77	70.08
45	11	68.33	68.49	68.68	68.48
44	12	66.88	67.02	67.44	66.90
43	13	65.43	65.55	66.07	65.35
42	20	63.98	64.07	63.81	63.82
41	22	62.52	62.60	61.40	62.30
40	16	61.07	61.13	59.54	60.80
39	14	59.62	59.66	58.29	59.31
38	10	58.17	58.19	57.05	57.84
37	5	56.71	56.72	55.91	56.37
36	12	55.26	55.25	54.61	54.91
35	9	53.81	53.78	53.41	53.45
34	16	52.36	52.30	51.65	52.00
33	6	50.91	50.83	50.38	50.56
32	9	49.45	49.36	49.26	49.12
31	6	48.00	47.89	47.89	47.68
30	8	46.55	46.42	46.77	46.24
29	5	45.10	44.95	45.65	44.81
28	7	43.65	43.48	44.67	43.37
27	7	42.19	42.00	43.34	41.94
26	5	40.74	40.53	41.79	40.50
25	6	39.29	39.06	40.57	39.07
24	9	37.84	37.59	39.03	37.64
23	8	36.39	36.12	36.90	36.20
22	2	34.93	34.65	35.53	34.76
21	3	33.48	33.18	34.74	33.32
20	5	32.03	31.71	33.52	31.88
19	3	30.58	30.23	30.71	30.43
18	2	29.12	28.76	29.28	28.98
17	5	27.67	27.29	27.51	27.52
16	2	26.22	25.82	26.02	26.06
15	3	24.77	24.35	24.02	24.59
14	0	23.32	22.88	22.68	23.11
13	2	21.86	21.41	22.01	21.63
12	3	20.41	19.94	20.51	20.13
11	1	18.96	18.46	18.51	18.62
10	0	17.51	16.99	18.01	17.09
9	1	16.06	15.52	17.01	15.54
8	0	14.60	14.05	16.01	13.98
7	0	13.15	12.58	16.01	12.38
6	1	11.70	11.11	14.01	10.76
5	0	10.25	9.64	13.00	9.10
4	0	8.80	8.16	13.00	7.41
3	0	7.34	6.69	13.00	5.66
2	0	5.89	5.22	13.00	3.84
1	0	4.44	3.75	13.00	1.97

TABLE 8

CAT14 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (819) - Grade 4, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT14 Converted Score LO	LT	EQ	RM
52	3	76.89	77.14	83.67	82.43
51	9	75.44	75.68	80.80	80.00
50	8	73.99	74.22	78.12	77.70
49	6	72.54	72.75	75.00	75.52
48	9	71.09	71.29	73.00	73.43
47	14	69.64	69.82	71.17	71.42
46	10	68.19	68.36	68.38	69.49
45	15	66.74	66.89	67.05	67.63
44	15	65.29	65.43	65.44	65.82
43	15	63.84	63.96	63.71	64.06
42	17	62.39	62.50	61.60	62.35
41	11	60.94	61.03	59.65	60.68
40	11	59.49	59.57	58.08	59.05
39	17	58.04	58.10	56.25	57.44
38	15	56.59	56.64	54.44	55.87
37	8	55.14	55.17	52.82	54.33
36	7	53.69	53.71	51.33	52.80
35	17	52.24	52.25	49.50	51.30
34	13	50.79	50.78	47.74	49.82
33	10	49.34	49.32	46.40	48.35
32	6	47.89	47.85	45.29	46.90
31	5	46.44	46.39	44.56	45.46
30	7	44.99	44.92	43.73	44.03
29	6	43.54	43.46	42.38	42.61
28	5	42.09	41.99	41.27	41.20
27	5	40.64	40.53	40.22	39.80
26	10	39.19	39.06	38.86	38.40
25	11	37.74	37.60	37.36	37.01
24	6	36.29	36.13	36.14	35.62
23	3	34.84	34.67	35.30	34.24
22	3	33.39	33.21	34.73	32.85
21	7	31.94	31.74	33.87	31.47
20	6	30.49	30.28	33.00	30.09
19	4	29.04	28.81	32.23	28.70
18	4	27.59	27.35	31.50	27.32
17	8	26.14	25.88	30.00	25.92
16	7	24.69	24.42	28.00	24.53
15	3	23.24	22.95	26.40	23.13
14	4	21.79	21.49	24.75	21.72
13	2	20.34	20.02	20.50	20.30
12	5	18.89	18.56	18.50	18.87
11	1	17.44	17.09	17.50	17.43
10	2	15.99	15.63	17.00	15.98
9	2	14.54	14.16	15.75	14.51
8	1	13.09	12.70	15.00	13.02
7	0	11.64	11.24	14.67	11.51
6	0	10.19	9.77	14.67	9.88
5	0	8.74	8.31	14.67	8.42
4	0	7.29	6.84	14.67	6.83
3	0	5.84	5.38	14.67	5.21
2	1	4.39	3.91	14.33	3.53
1	0	2.94	2.45	14.00	1.80

CAT15 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (820) - Grade 5, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT15 Converted Score LO	LT	EQ	RM
51	0	86.08	86.73	83.00	83.81
50	0	84.53	85.14	83.00	82.58
49	0	82.97	83.56	83.00	81.30
48	5	81.42	81.97	80.50	79.99
47	3	79.86	80.39	77.43	78.65
46	3	78.30	78.80	76.67	77.29
45	6	76.75	77.22	75.67	75.91
44	11	75.19	75.63	73.82	74.50
43	5	73.64	74.05	72.42	73.09
42	10	72.08	72.46	71.17	71.67
41	10	70.53	70.88	69.67	70.24
40	11	68.97	69.29	68.53	68.79
39	13	67.41	67.71	67.22	67.34
38	10	65.86	66.12	65.83	65.87
37	16	64.30	64.54	64.50	64.40
36	8	62.75	62.95	63.19	62.92
35	12	61.19	61.36	61.45	61.43
34	16	59.63	59.78	59.38	59.94
33	11	58.08	58.19	57.41	58.43
32	13	56.52	56.61	56.19	56.92
31	13	54.97	55.02	54.91	55.39
30	8	53.41	53.44	53.96	53.86
29	11	51.86	51.85	53.17	52.32
28	13	50.30	50.27	51.50	50.77
27	10	48.74	48.68	50.14	49.21
26	10	47.19	47.10	49.23	47.64
25	13	45.63	45.51	47.76	46.05
24	17	44.08	43.93	45.73	44.46
23	6	42.52	42.34	43.33	42.85
22	11	40.97	40.76	41.71	41.23
21	7	39.41	39.17	40.57	39.60
20	6	37.85	37.59	39.62	37.96
19	15	36.30	36.00	38.07	36.30
18	11	34.74	34.42	34.93	34.63
17	9	33.19	32.83	33.22	32.94
16	12	31.63	31.25	30.14	31.24
15	7	30.07	29.66	27.60	29.51
14	3	28.52	28.08	25.86	27.77
13	3	26.96	26.49	25.00	26.01
12	7	25.41	24.91	23.60	24.22
11	5	23.85	23.32	20.91	22.42
10	5	22.30	21.74	20.00	20.59
9	2	20.74	20.15	19.12	18.73
8	2	19.18	18.57	16.50	16.84
7	1	17.63	16.98	13.00	14.91
6	0	16.07	15.40	12.00	12.95
5	0	14.52	13.81	12.00	10.95
4	0	12.96	12.23	12.00	8.90
3	0	11.40	10.64	12.00	6.80
2	0	9.85	9.06	12.00	4.62
1	0	8.29	7.47	12.00	2.37

TABLE 10

CAT16 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (820) - Grade 6, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model).

ITEM BANK RAW SCORE	FREQUENCY	CAT16 Converted Score LO	LT	EQ	RM
47	2	75.05	75.42	80.00	81.95
46	8	73.41	73.75	78.00	79.11
45	13	71.77	72.08	76.44	76.42
44	16	70.13	70.42	72.68	73.88
43	16	68.49	68.75	70.08	71.45
42	20	66.85	67.08	67.94	69.11
41	22	65.21	65.42	65.98	66.86
40	25	63.57	63.75	63.22	64.70
39	20	61.92	62.08	61.12	62.59
38	23	60.28	60.42	59.13	60.55
37	18	58.64	58.75	57.80	58.56
36	24	57.00	57.08	55.89	56.62
35	17	55.36	55.42	53.12	54.72
34	15	53.72	53.75	51.47	52.87
33	16	52.08	52.08	50.04	51.05
32	19	50.44	50.42	48.40	49.26
31	11	48.80	48.75	46.29	47.50
30	11	47.16	47.08	45.09	45.77
29	8	45.52	45.42	44.17	44.06
28	8	43.88	43.75	43.27	42.38
27	15	42.24	42.08	41.75	40.72
26	6	40.60	40.42	40.04	39.07
25	9	38.96	38.75	38.65	37.44
24	11	37.32	37.08	37.49	35.83
23	7	35.68	35.41	36.13	34.23
22	8	34.04	33.75	34.83	32.64
21	10	32.40	32.08	33.74	31.07
20	11	30.76	30.41	32.35	29.51
19	2	29.12	28.75	30.70	27.95
18	9	27.48	27.08	29.02	26.41
17	6	25.84	25.41	27.02	24.87
16	5	24.20	23.75	25.68	23.35
15	3	22.56	22.08	24.79	21.83
14	4	20.92	20.41	24.01	20.31
13	2	19.28	18.75	23.01	18.81
12	0	17.64	17.08	22.51	17.31
11	2	16.00	15.41	22.01	15.82
10	3	14.36	13.75	19.02	14.33
9	2	12.72	12.08	17.00	12.85
8	1	11.08	10.41	15.00	11.38
7	0	9.44	8.75	14.00	9.92
6	0	7.80	7.08	14.00	8.46
5	0	6.16	5.41	14.00	7.02
4	0	4.52	3.75	14.00	5.58
3	0	2.88	2.08	14.00	4.16
2	0	1.24	.41	14.00	2.76
1	0	-.40	-1.25	14.00	1.37

TABLE 11

CAT17 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (821) - Grade 7, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT17 Converted Score LO	LT	EQ	RM
48	4	79.21	79.77	83.00	82.66
47	5	77.58	78.11	80.64	80.40
46	8	75.96	76.46	78.63	78.22
45	7	74.33	74.80	77.58	76.11
44	10	72.71	73.15	76.51	74.06
43	13	71.08	71.49	75.31	72.06
42	15	69.46	69.84	73.69	70.12
41	18	67.83	68.19	70.69	68.21
40	21	66.21	66.53	67.24	66.35
39	15	64.58	64.88	65.32	64.52
38	15	62.96	63.22	62.51	62.73
37	20	61.33	61.57	60.25	60.96
36	15	59.71	59.91	57.96	59.22
35	14	58.09	58.26	56.37	57.52
34	22	56.46	56.60	54.20	55.83
33	16	54.84	54.95	52.41	54.16
32	20	53.21	53.30	50.86	52.52
31	25	51.59	51.64	49.05	50.89
30	20	49.96	49.99	47.24	49.27
29	13	48.34	48.33	46.12	47.67
28	14	46.71	46.68	45.17	46.09
27	15	45.09	45.02	43.90	44.51
26	13	43.46	43.37	42.79	42.95
25	16	41.84	41.71	41.70	41.40
24	19	40.21	40.06	40.10	39.85
23	16	38.59	38.41	38.10	38.31
22	12	36.96	36.75	36.65	36.78
21	16	35.34	35.10	35.52	35.25
20	10	33.71	33.44	34.37	33.72
19	20	32.09	31.79	32.76	32.20
18	12	30.47	30.13	30.77	30.67
17	16	28.84	28.48	29.18	29.14
16	9	27.22	26.82	28.24	27.62
15	8	25.59	25.17	27.51	26.08
14	8	23.97	23.52	26.66	24.54
13	7	22.34	21.86	25.62	22.99
12	10	20.72	20.21	23.78	21.43
11	12	19.09	18.55	22.02	19.86
10	5	17.47	16.90	20.24	18.27
9	6	15.84	15.24	18.42	16.66
8	0	14.22	13.59	17.21	15.03
7	3	12.59	11.93	16.02	13.36
6	0	10.97	10.28	13.01	11.66
5	1	9.34	8.63	12.51	9.92
4	0	7.72	6.97	12.00	8.12
3	0	6.10	5.32	12.00	6.26
2	0	4.47	3.66	12.00	4.30
1	0	2.85	2.01	12.00	2.23

TABLE 12

CAT18 Converted Total Mathematics Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (821) - Grade 8, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	CAT18 Converted Score LO	LT	EQ	RM
46	14	80.39	80.92	82.35	82.51
45	11	78.63	79.12	79.98	80.15
44	22	76.88	77.33	78.29	77.89
43	28	75.12	75.54	76.34	75.71
42	26	73.36	73.74	73.75	73.61
41	22	71.61	71.95	71.24	71.57
40	17	69.85	70.16	69.48	69.59
39	19	68.09	68.36	67.78	67.65
38	16	66.34	66.57	65.96	65.75
37	24	64.58	64.78	64.03	63.89
36	15	62.82	62.99	62.29	62.07
35	27	61.07	61.19	60.82	60.28
34	20	59.31	59.40	58.72	58.51
33	19	57.55	57.61	57.04	56.77
32	23	55.80	55.81	55.26	55.04
31	13	54.04	54.02	53.92	53.34
30	19	52.28	52.23	52.19	51.65
29	15	50.53	50.43	49.99	49.97
28	13	48.77	48.64	48.61	48.31
27	15	47.01	46.85	47.02	46.65
26	13	45.26	45.05	45.40	45.01
25	8	43.50	43.26	44.57	43.37
24	14	41.74	41.47	43.25	41.74
23	19	39.99	39.67	41.47	40.12
22	15	38.23	37.88	39.75	38.49
21	11	36.47	36.09	38.11	36.87
20	11	34.72	34.30	36.42	35.24
19	14	32.96	32.50	34.59	33.62
18	22	31.20	30.71	32.25	31.99
17	8	29.45	28.92	29.82	30.36
16	8	27.69	27.12	28.09	28.72
15	6	25.93	25.33	25.69	27.07
14	10	24.18	23.54	23.49	25.42
13	6	22.42	21.74	19.24	23.75
12	4	20.66	19.95	16.45	22.07
11	1	18.91	18.16	15.53	20.38
10	1	17.15	16.36	15.02	18.67
9	1	15.39	14.57	14.03	16.94
8	2	13.64	12.78	12.62	15.19
7	0	11.88	10.98	11.02	13.42
6	0	10.12	9.19	11.02	11.62
5	0	8.37	7.40	11.02	9.79
4	1	6.61	5.61	7.00	7.93
3	0	4.85	3.81	6.00	6.02
2	0	3.10	2.02	6.00	4.07
1	0	1.34	.23	6.00	2.07

APPENDIX C

Graphs of CAT/C Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test

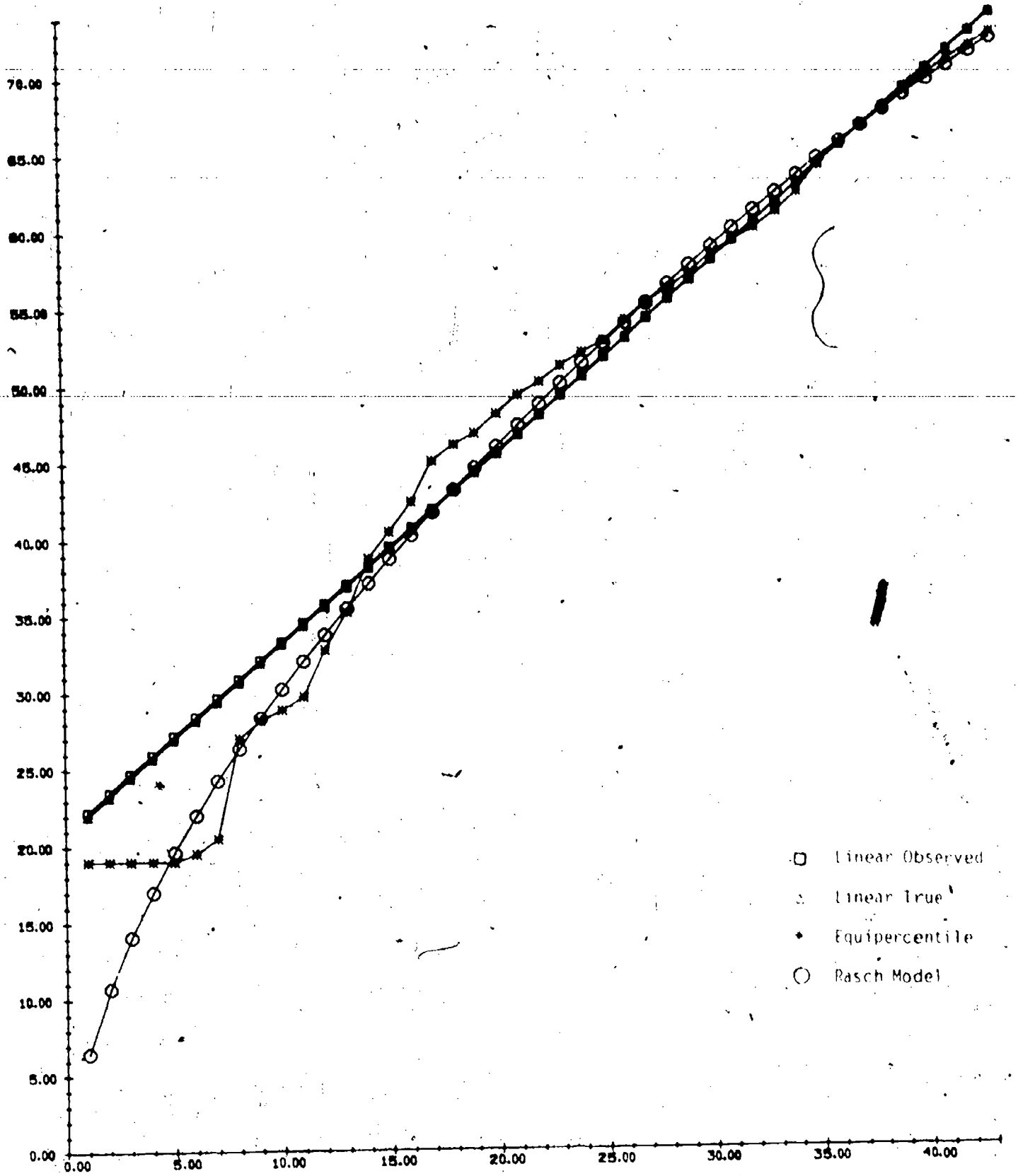

FIGURE 1

Graph of CAT13 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (318) - Grade 3, Reading.

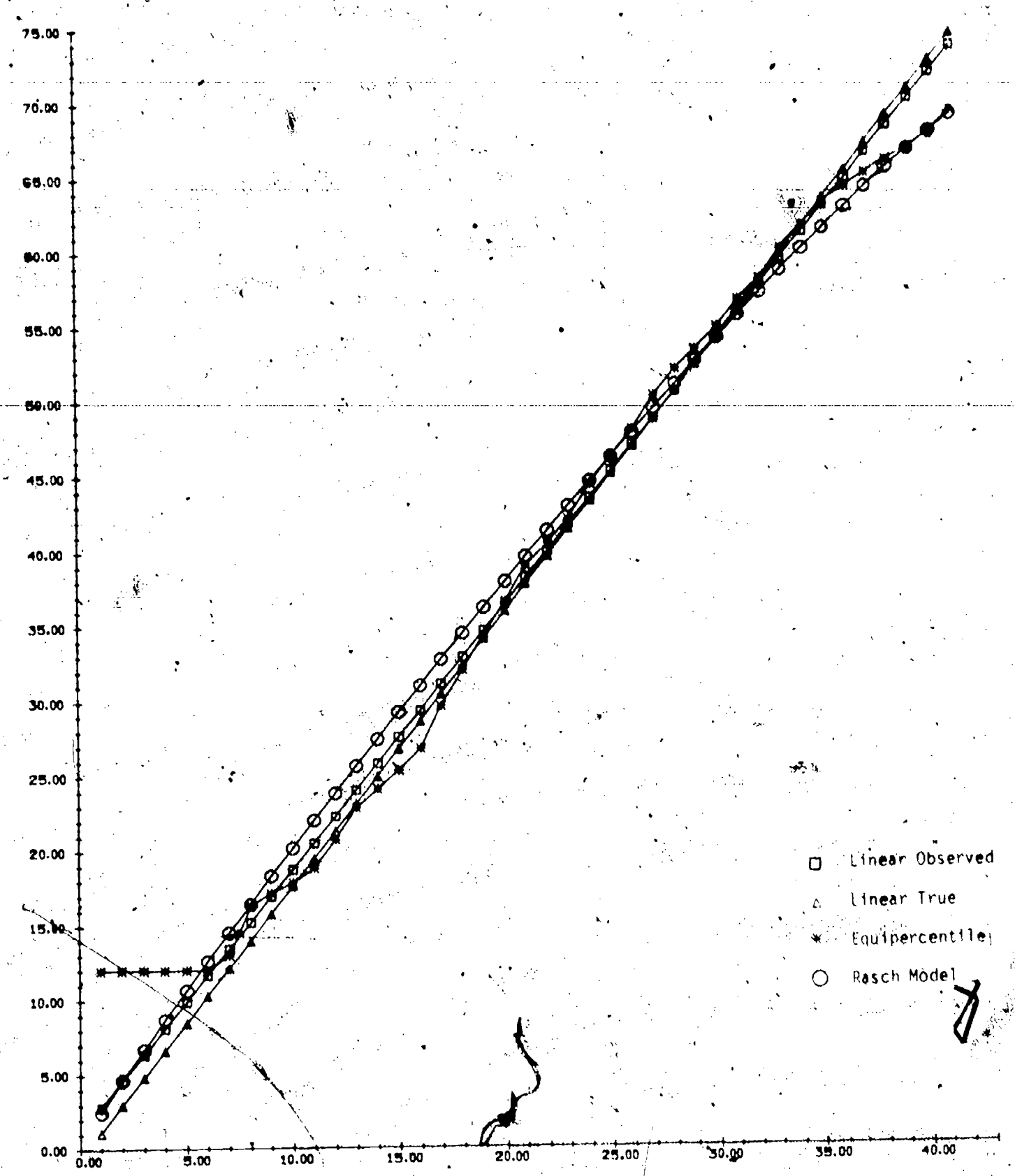

FIGURE 2

Graph of CAT14 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (319) - Grade 4, Reading.

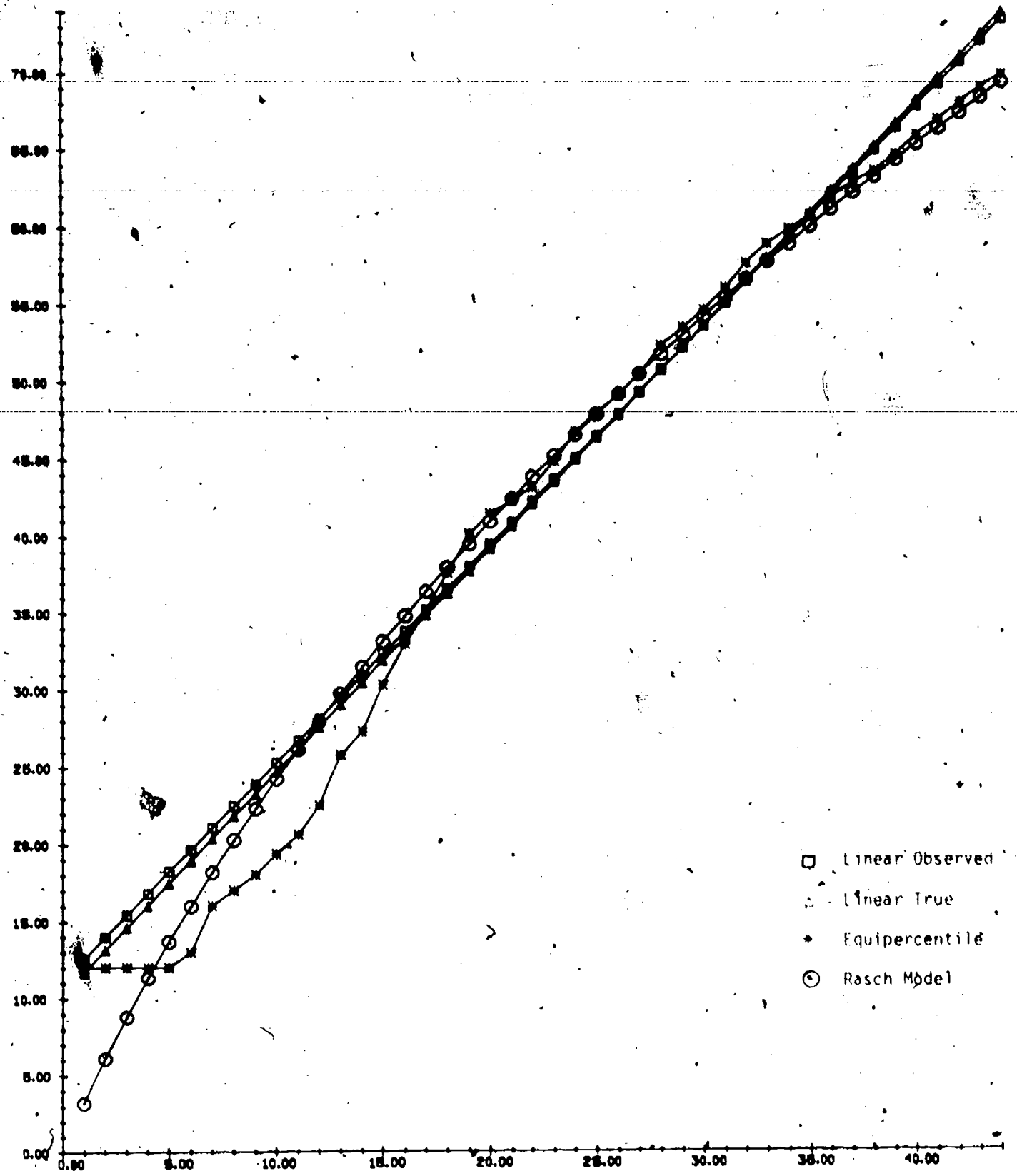

FIGURE 3

Graph of CAT15 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (320) - Grade 5, Reading.

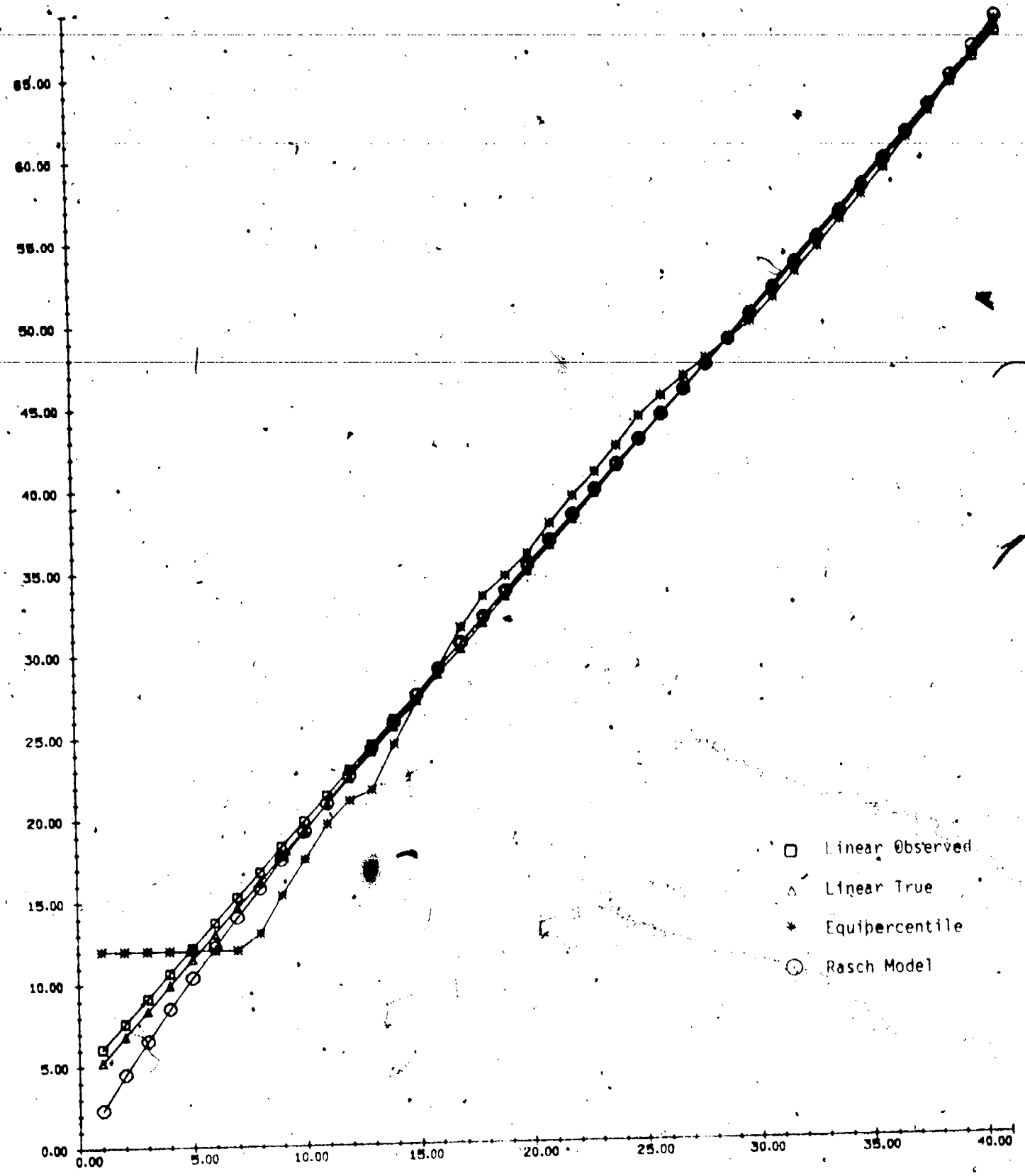

FIGURE 4

Graph of CAT16 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (320) - Grade 6, Reading.

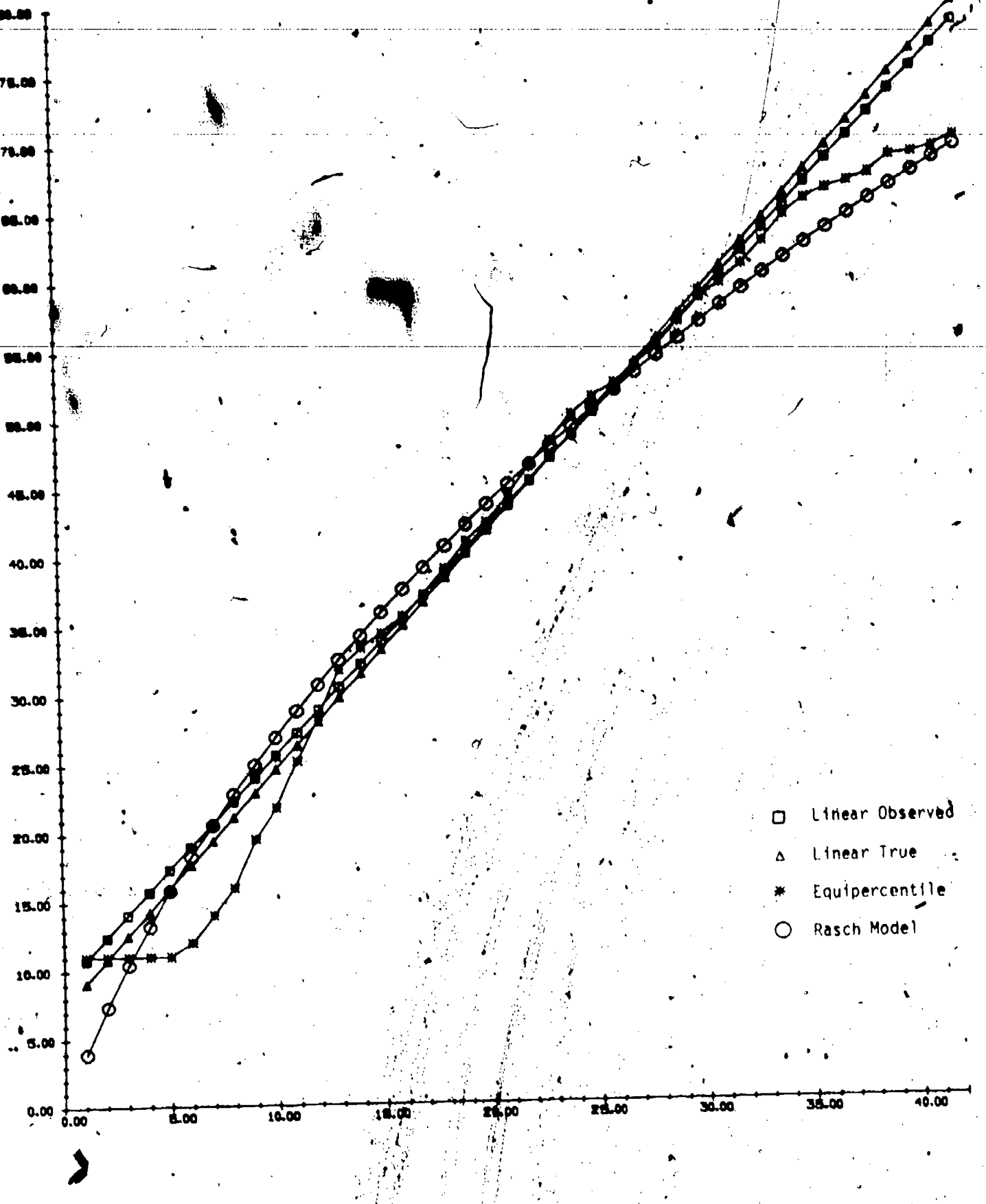

FIGURE 5

Graph of CAT17 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (321) - Grade 7, Reading.

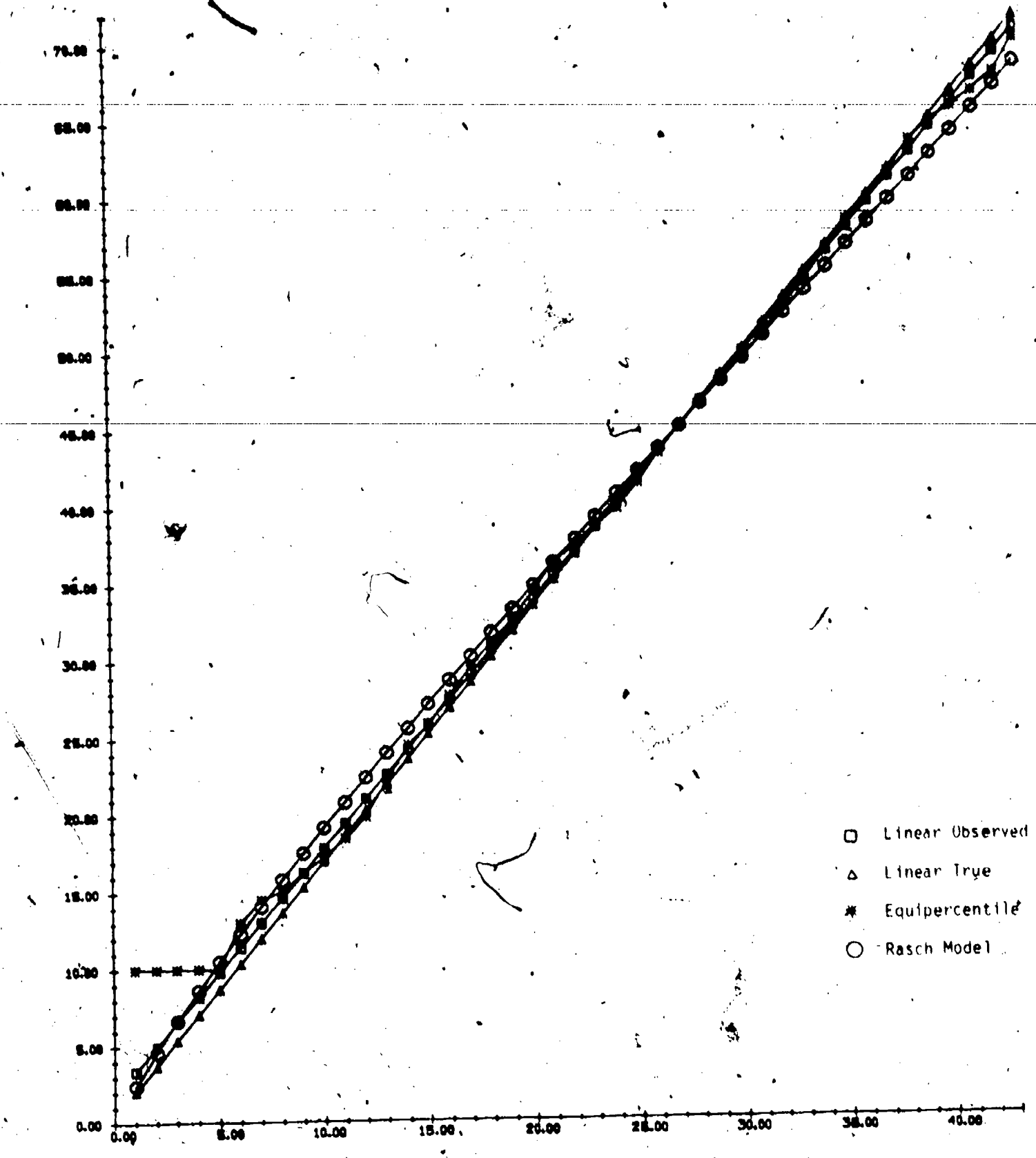

FIGURE 5

Graph of CAT18 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (321), - Grade 8, Reading.

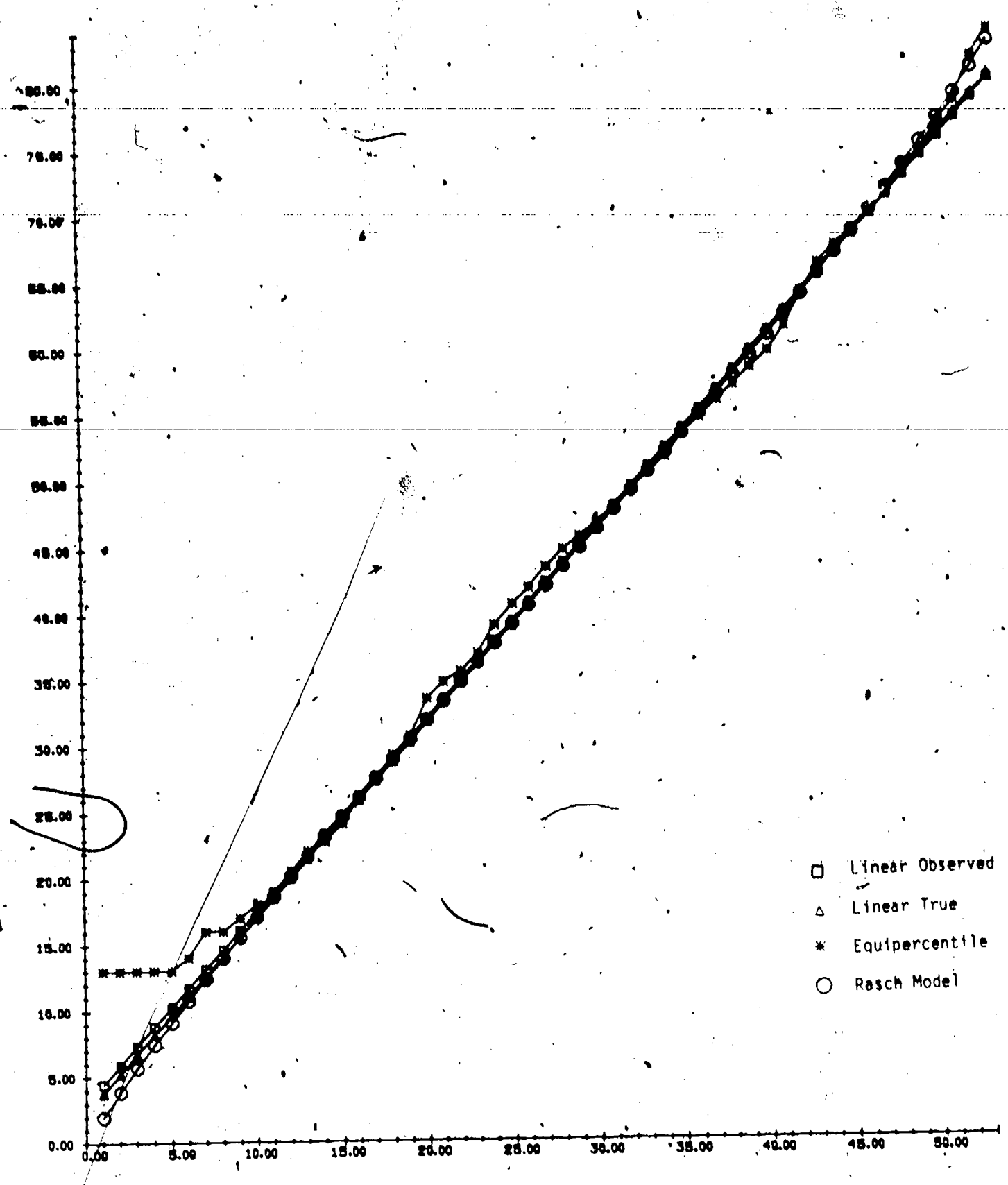

FIGURE 7

Graph of CAT13.Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (818) - Grade 3, Mathematics.

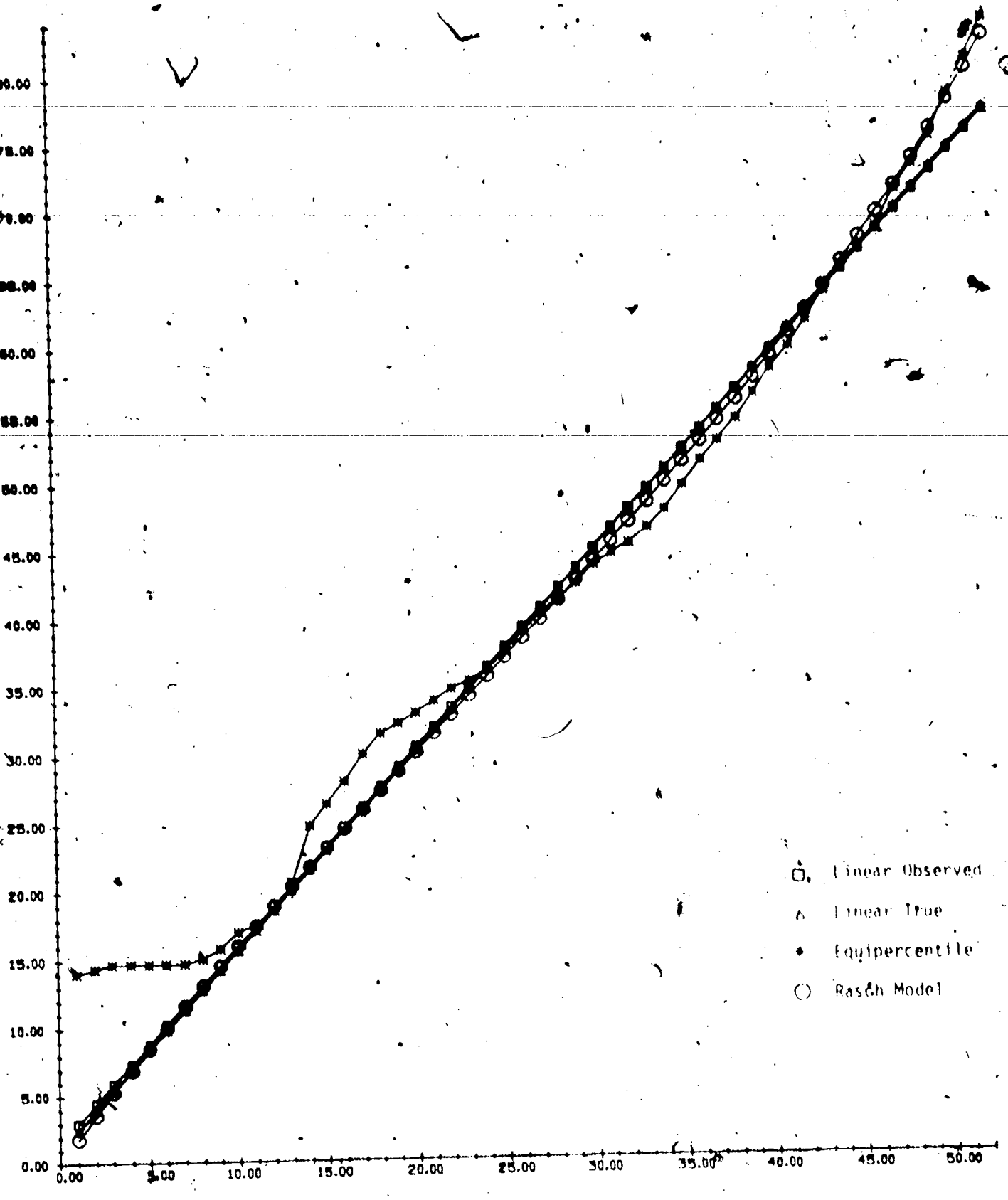

FIGURE 8

Graph of CAT14 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (8T9) - Grade 4, Mathematics.

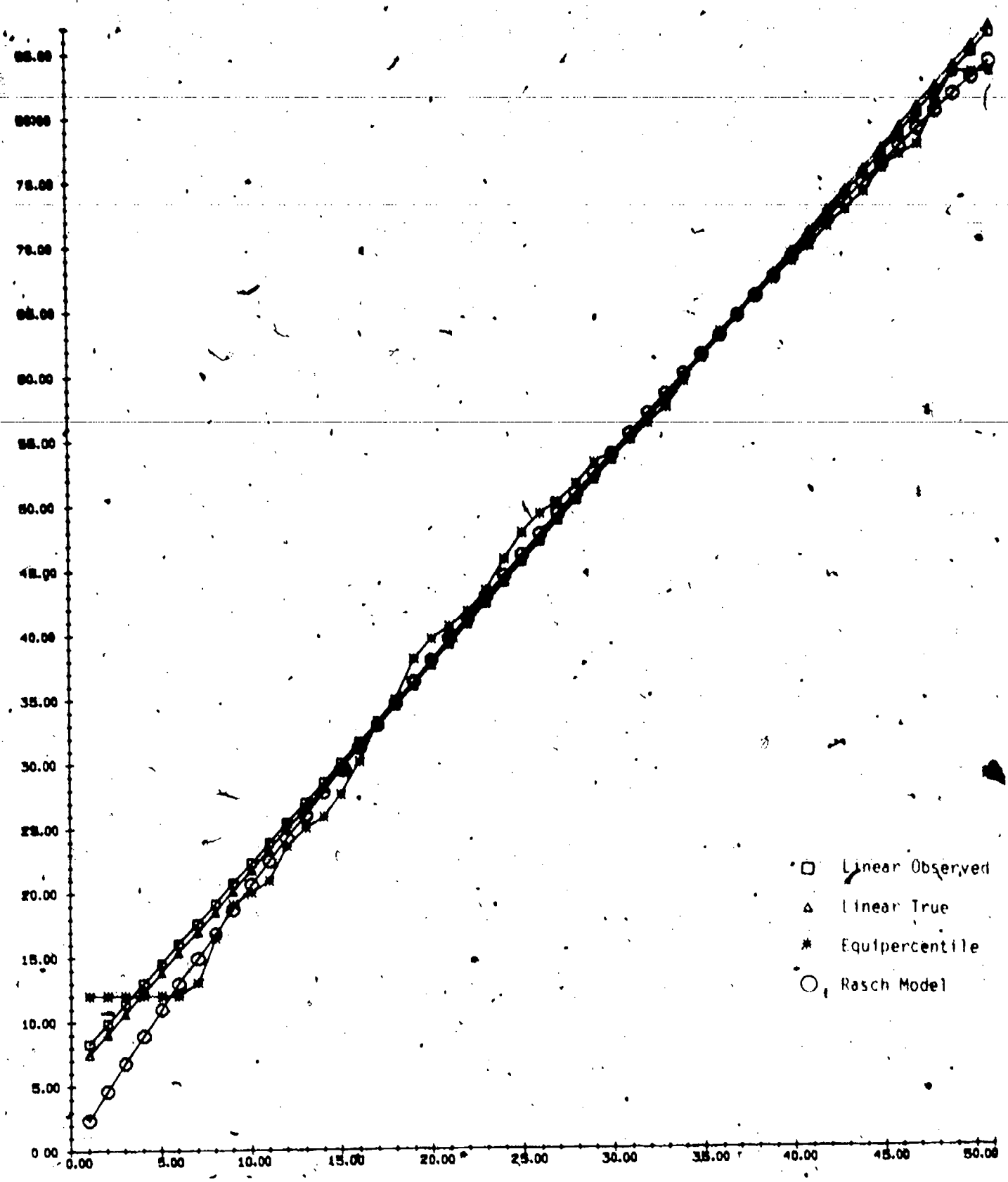

FIGURE 9

Graph of CAT15 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (820) - Grade 5, Mathematics.

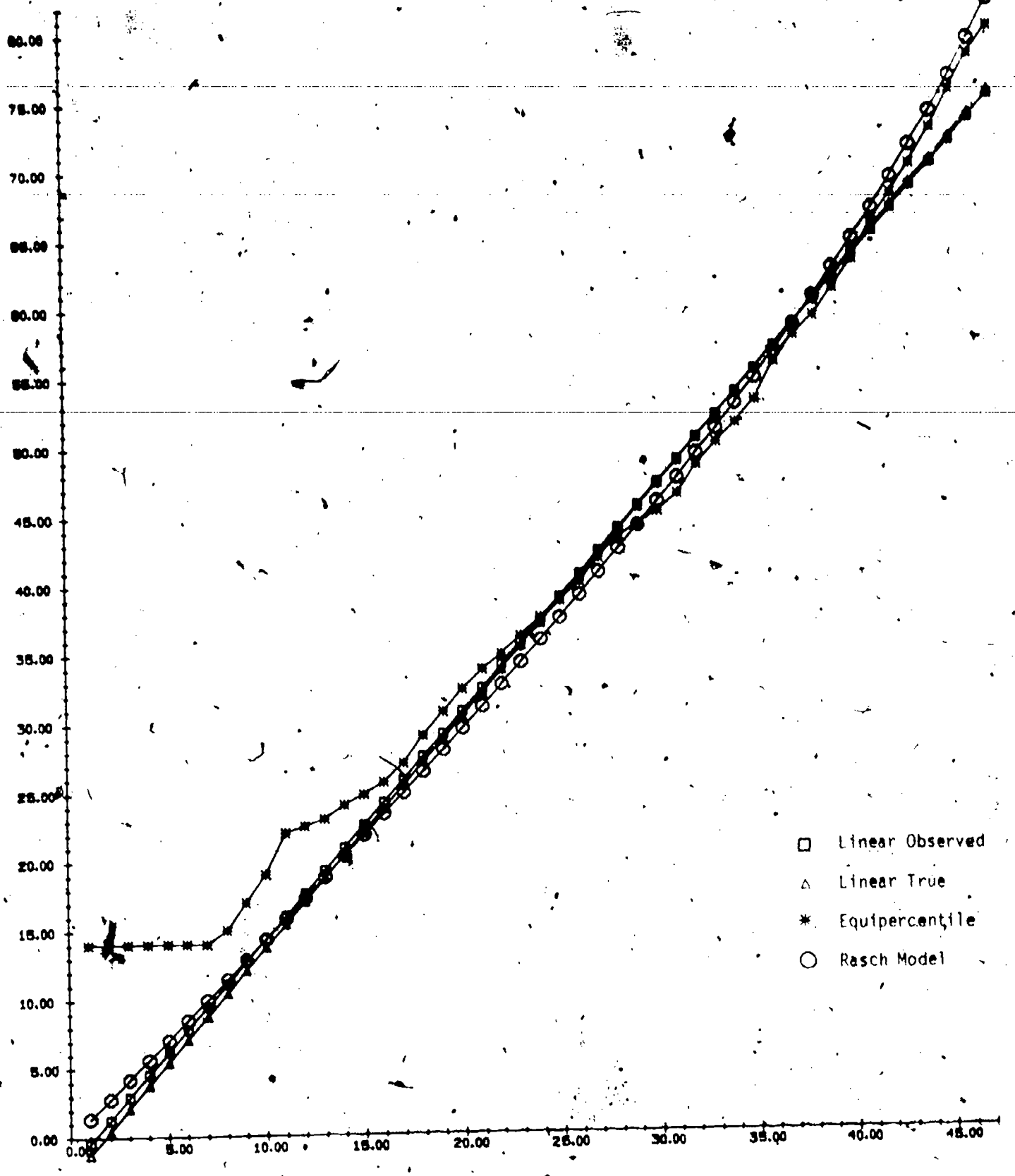

FIGURE 10

Graph of CAT10 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (820) - Grade 6, Mathematics.

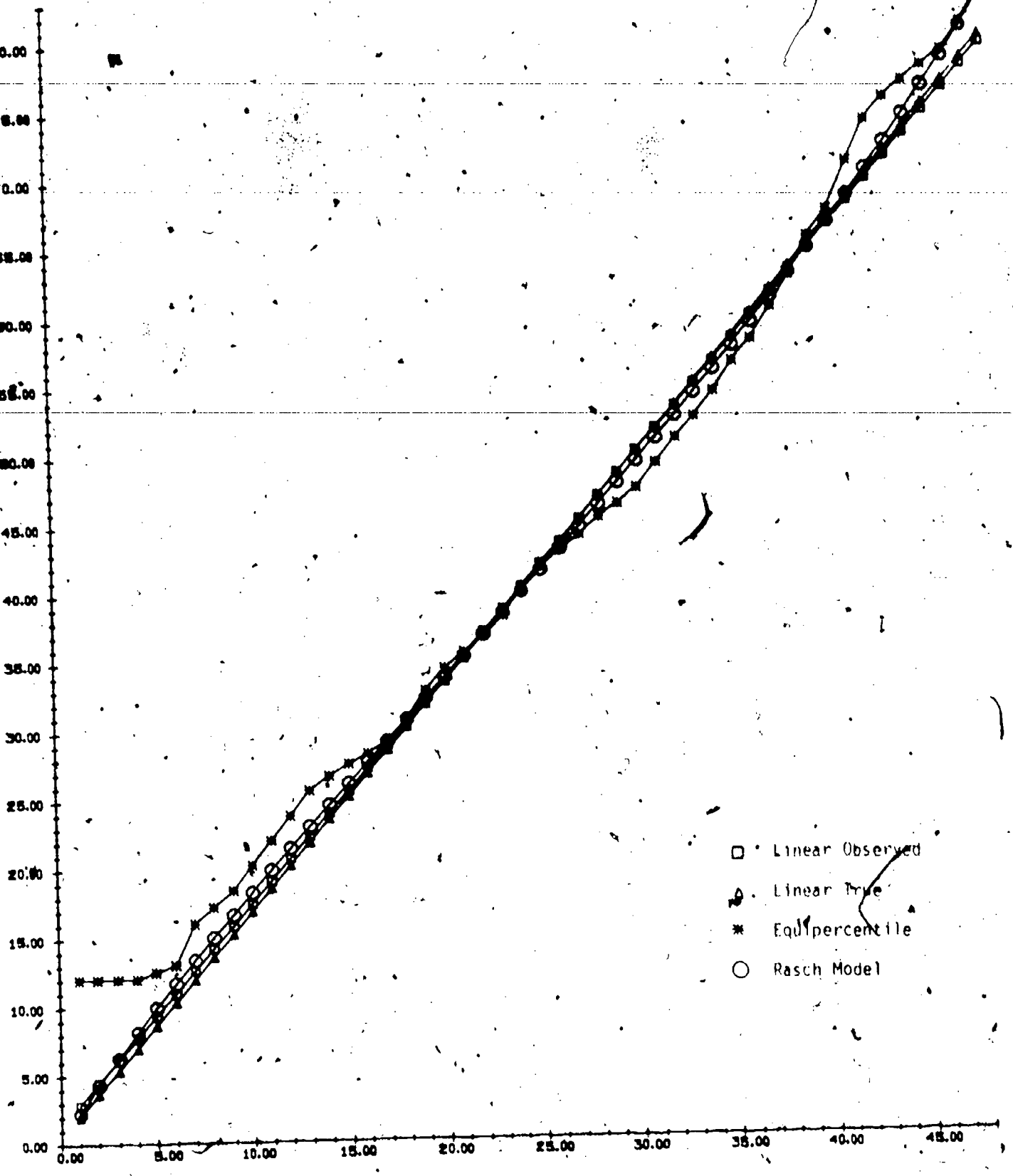

FIGURE 11

Graph of CAT17 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (821) - Grade 7, Mathematics.

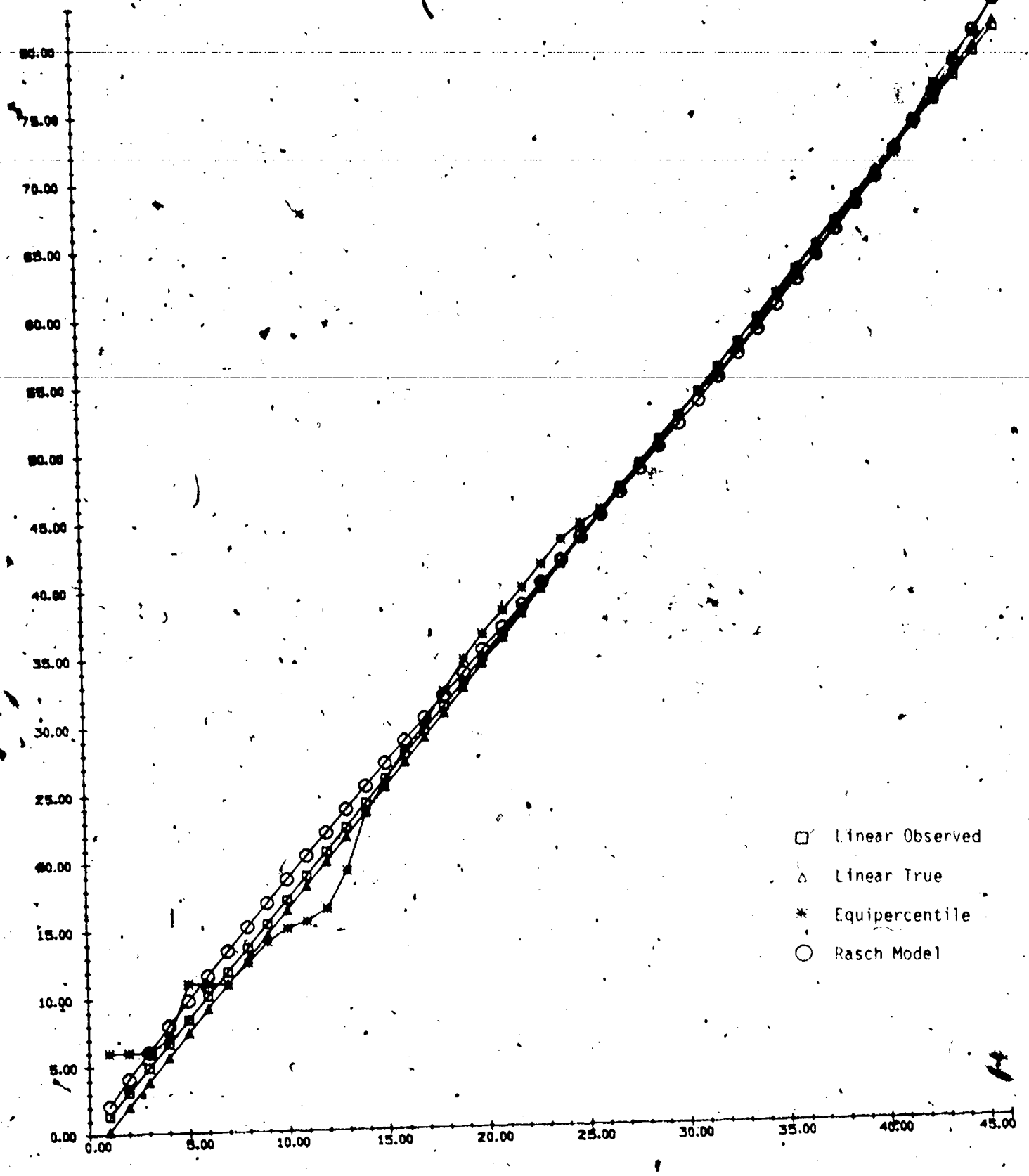

FIGURE 12

Graph of CAT18 Converted Scores Yielded by Four Equating Methods for Scores on the Corresponding Item Bank Test (821) - Grade 8, Mathematics.

APPENDIX D

Average Absolute Discrepancy Between CAT/C Converted Scores Yielded by Four Equating Techniques and Those Actually Observed at Each Item Bank Raw Score

TABLE 1

Average Discrepancy Between CAT13 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (318) - Grade 3, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
43	7	2.86	2.97	1.67	1.41
42	16	2.79	2.86	2.15	1.96
41	24	2.99	3.04	2.72	2.65
40	15	3.19	3.22	3.08	2.96
39	29	3.03	3.04	3.00	2.98
38	29	4.14	4.16	4.07	4.06
37	26	3.73	3.74	3.73	3.73
36	26	3.48	3.47	3.45	3.43
35	27	4.68	4.71	4.66	4.89
34	30	5.15	5.15	5.31	5.08
33	16	7.24	7.24	7.32	7.21
32	27	4.73	4.73	4.66	4.88
31	7	3.69	3.69	3.68	4.49
30	20	4.83	4.84	4.74	4.65
29	17	6.31	6.31	6.44	6.64
28	12	5.81	5.80	6.03	6.18
27	11	5.40	5.41	5.13	5.15
26	12	7.25	7.25	7.25	7.25
25	6	10.01	10.04	9.33	9.42
24	3	7.18	7.18	6.66	6.89
23	9	11.66	11.72	10.76	11.10
22	6	6.81	6.84	6.50	6.60
21	5	8.37	8.35	8.87	8.47
20	9	10.98	11.06	9.27	10.68
19	5	13.98	14.04	12.48	13.85
18	3	11.29	11.32	10.38	11.29
17	5	19.22	19.24	18.59	19.26
16	9	9.12	9.07	9.68	8.95
15	8	8.61	8.58	8.86	8.41
14	8	5.93	5.90	6.04	5.88
13	8	12.00	12.00	12.00	12.00
12	8	16.18	16.23	17.30	16.75
11	4	15.21	15.39	20.05	17.74
10	2	28.69	28.87	33.18	31.80
9	2	14.42	14.61	18.34	18.18
8	3	13.48	13.68	17.34	18.00
7	1	4.38	4.59	13.51	9.76
6	1	12.60	12.82	21.50	18.99

TABLE 2

Average Discrepancy Between CAT14 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (319) - Grade 4, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY			
		LO	LT	EQ	RM
41	3	5.87	6.74	2.08	1.98
40	11	5.68	6.49	1.96	1.98
39	12	3.68	4.42	1.33	1.33
38	14	3.95	4.62	1.99	1.92
37	15	5.48	5.94	4.58	4.24
36	13	3.48	3.90	3.14	2.75
35	17	3.69	3.78	3.73	3.50
34	28	5.15	5.27	5.30	5.04
33	26	4.65	4.65	4.66	4.65
32	19	5.16	5.12	5.09	5.24
31	23	4.63	4.60	4.58	4.66
30	20	4.97	5.00	5.12	4.97
29	20	6.20	6.20	6.24	6.20
28	22	6.22	6.22	6.10	6.18
27	21	6.34	6.34	6.10	6.22
26	27	7.34	7.32	7.46	7.43
25	12	5.88	5.96	5.47	5.46
24	20	8.40	8.42	8.31	8.29
23	16	7.57	7.64	7.49	7.27
22	11	9.04	9.26	8.80	8.42
21	14	10.00	9.83	10.34	10.61
20	13	9.65	10.06	9.50	8.53
19	10	5.70	5.70	5.70	5.74
18	13	9.07	9.02	9.00	9.19
17	11	9.94	9.75	9.54	10.38
16	7	14.52	14.63	14.94	14.29
15	10	4.71	4.60	4.74	5.28
14	6	6.57	6.85	7.14	6.17
13	6	5.67	5.67	5.67	5.67
12	8	3.25	3.25	3.32	3.46
11	3	10.50	11.23	11.88	10.00
10	2	10.00	10.39	10.17	10.00
9	2	.05	1.23	.16	1.29
8	1	2.82	4.06	1.67	1.59
7	1	9.58	10.89	10.01	8.48

TABLE 3

Average Discrepancy Between CAT15 Converted Scores Yielded By Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (320) - Grade 5, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
44	4	4.68	5.16	1.28	1.03
43	5	6.57	7.02	3.75	3.62
42	12	5.11	5.53	3.30	2.97
41	10	2.16	2.54	1.26	1.38
40	21	5.39	5.72	3.93	3.62
39	20	3.18	3.37	2.75	2.71
38	18	2.50	2.50	2.70	2.82
37	13	4.78	4.92	4.53	4.38
36	24	4.58	4.72	4.49	4.19
35	16	3.63	3.63	3.63	3.67
34	13	3.78	3.79	3.82	3.79
33	22	3.60	3.54	3.20	3.65
32	17	4.87	4.86	4.85	4.86
31	23	6.59	6.60	6.80	6.67
30	20	4.26	4.25	4.15	4.15
29	20	5.95	5.94	5.89	5.86
28	21	6.35	6.36	5.72	5.92
27	18	4.90	4.91	4.65	4.64
26	14	3.98	4.01	3.93	3.93
25	14	5.11	5.17	4.60	4.63
24	18	4.83	4.83	5.20	5.16
23	17	6.43	6.46	6.22	6.18
22	13	8.45	8.47	8.38	8.33
21	6	5.41	5.65	4.00	3.86
20	13	6.08	6.02	6.97	6.70
19	15	8.22	8.37	7.52	7.74
18	12	5.49	5.55	5.33	5.27
17	10	11.75	11.90	11.76	11.34
16	12	8.42	8.42	8.43	8.42
15	9	6.59	6.57	6.75	6.70
14	7	6.76	6.96	8.31	6.53
13	9	8.84	9.00	10.14	8.77
12	4	6.71	6.98	11.78	6.79
11	3	3.67	3.11	2.41	3.10
10	2	1.24	1.83	7.17	2.29
9	2	3.85	3.23	2.01	2.26
8	1	13.56	14.21	19.01	15.76
6	1	12.37	13.09	19.01	16.04

TABLE 4

Average Discrepancy Between CAT16 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (320) - Grade 6, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
41	8	3.13	3.13	3.13	3.14
40	15	2.35	2.50	2.49	2.73
39	23	3.00	3.17	3.07	3.22
38	20	3.45	3.48	3.44	3.48
37	22	3.19	3.28	3.14	3.22
36	25	5.14	5.23	4.91	5.13
35	27	4.25	4.31	4.19	4.25
34	21	4.76	4.82	4.55	4.70
33	21	4.12	4.09	4.28	4.15
32	16	4.77	4.81	4.56	4.71
31	29	5.96	5.98	5.74	5.91
30	16	5.46	5.45	5.57	5.48
29	18	4.73	4.73	4.77	4.77
28	14	4.83	4.83	4.95	4.80
27	11	6.52	6.52	6.59	6.52
26	12	5.42	5.42	5.47	5.42
25	12	4.60	4.64	4.50	4.60
24	22	6.06	6.09	5.91	6.05
23	12	6.37	6.42	6.25	6.35
22	14	7.25	7.31	6.95	7.23
21	15	7.24	7.22	7.31	7.24
20	11	6.06	6.21	5.70	6.03
19	8	10.72	10.94	9.98	10.67
18	11	7.36	7.39	7.30	7.36
17	9	5.72	5.60	6.13	5.73
16	10	6.90	6.90	6.90	6.90
15	9	4.19	4.32	4.33	4.21
14	12	10.20	10.34	10.70	10.25
13	2	10.12	10.58	12.88	10.36
12	3	3.72	3.88	5.00	3.83
11	4	11.00	11.00	11.00	11.00
10	4	4.75	4.75	5.50	4.75
9	2	2.00	2.31	4.67	2.52
8	2	3.76	4.37	7.50	4.75

TABLE 5

Average Discrepancy Between CAT17 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (321) - Grade 7, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
42	1	8.14	9.60	.25	.89
41	2	6.99	8.38	.50	1.31
39	1	7.21	8.44	2.41	.29
38	7	7.70	8.86	4.07	3.76
37	2	5.42	6.50	2.09	.50
36	7	6.71	7.71	4.55	3.26
35	10	6.08	6.75	5.81	6.45
34	15	3.46	3.75	3.32	3.95
33	12	4.84	5.23	4.36	3.83
32	17	4.39	4.48	4.37	4.60
31	11	5.35	5.18	5.65	6.36
30	16	5.05	5.12	5.04	5.00
29	23	4.84	5.01	4.81	4.59
28	20	3.63	3.70	3.61	3.68
27	21	6.93	7.03	6.95	6.77
26	17	5.69	5.62	5.56	5.72
25	13	6.92	6.98	7.41	7.03
24	27	5.57	5.57	5.68	5.57
23	16	8.00	8.00	8.00	8.00
22	30	6.10	6.11	5.97	5.97
21	29	6.55	6.53	6.65	6.91
20	18	5.29	5.38	5.22	5.06
19	16	7.32	7.09	7.75	8.67
18	25	7.32	7.43	7.25	6.78
17	19	8.34	8.61	8.34	7.51
16	21	7.81	8.26	7.70	6.50
15	9	8.25	8.32	8.19	8.21
14	14	6.27	6.53	6.14	6.17
13	15	4.70	4.72	4.78	4.83
12	14	7.14	7.50	7.35	6.36
11	13	5.42	5.63	6.03	5.05
10	9	8.71	9.04	10.24	8.27
9	3	3.30	2.94	3.19	3.62
8	2	9.74	9.89	15.06	8.24
7	3	6.79	7.20	9.01	6.81
6	1	16.03	17.33	23.01	16.74

TABLE 6

Average Discrepancy Between CAT18 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (321) - Grade 8, Reading. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
42	5	4.84	5.17	3.17	3.00
41	4	5.79	6.56	4.85	3.77
40	8	7.30	8.03	6.98	5.66
39	7	5.19	5.29	5.24	5.06
38	15	4.18	4.27	4.30	4.19
37	22	4.31	4.46	4.39	4.07
36	19	5.03	5.17	5.09	4.85
35	25	6.83	6.88	6.88	6.73
34	23	3.96	3.91	3.93	4.16
33	25	5.73	5.72	5.74	5.78
32	19	5.99	5.98	6.01	6.03
31	25	7.14	7.17	7.08	7.08
30	18	6.10	6.12	6.04	6.05
29	19	4.86	4.87	4.85	4.84
28	25	6.25	6.25	6.26	6.26
27	22	5.19	5.19	5.19	5.19
26	26	5.46	5.46	5.46	5.46
25	29	4.72	4.74	4.84	4.68
24	13	8.22	8.19	8.09	8.32
23	18	6.83	6.78	6.81	7.02
22	15	5.71	5.81	5.55	5.38
21	15	6.97	6.99	6.91	6.92
20	24	4.44	4.49	4.45	4.48
19	18	6.44	6.80	6.14	5.60
18	18	6.16	6.02	6.29	6.53
17	18	6.21	6.37	6.04	5.81
16	19	5.61	5.70	5.58	5.56
15	14	5.60	5.93	5.66	4.84
14	12	3.55	3.87	3.42	3.17
13	19	6.40	6.66	6.58	6.10
12	7	3.85	4.37	4.70	3.37
11	6	6.86	7.13	7.19	6.40
10	2	3.69	4.54	4.33	2.34
9	2	.71	.50	.75	1.99
8	1	31.40	32.35	31.00	30.21
6	2	4.50	5.16	4.50	4.50

TABLE 7

Average Discrepancy Between CAT13 Converted Scores Yielded By Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (818) - Grade 3, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
53	3	2.95	3.26	6.66	5.80
52	4	4.50	4.50	4.50	4.50
51	7	7.89	8.09	8.74	9.13
50	11	3.96	3.99	4.17	4.25
49	8	4.29	4.34	4.37	4.53
48	12	4.17	4.17	4.17	4.17
47	14	3.57	3.62	3.68	3.71
46	13	6.52	6.54	6.52	6.58
45	11	4.79	4.86	4.95	4.85
44	12	4.92	4.92	4.92	4.92
43	13	5.05	5.03	4.91	5.07
42	20	5.60	5.61	5.58	5.58
41	22	6.51	6.53	6.20	6.45
40	16	3.09	3.11	3.12	3.06
39	14	6.98	7.00	6.61	6.85
38	10	4.63	4.64	4.79	4.63
37	5	4.54	4.54	4.38	4.47
36	12	6.62	6.63	6.80	6.70
35	9	5.05	5.04	4.91	4.93
34	16	7.32	7.30	7.06	7.19
33	6	2.33	2.33	2.33	2.33
32	9	6.07	6.10	6.14	6.18
31	6	3.50	3.50	3.50	3.50
30	8	10.38	10.38	10.38	10.38
29	5	5.74	5.83	5.41	5.91
28	7	4.38	4.35	4.72	4.34
27	7	7.20	7.29	6.71	7.31
26	5	5.96	6.08	5.64	6.10
25	6	3.83	3.83	4.02	3.83
24	9	7.32	7.29	7.45	7.29
23	8	5.47	5.41	5.60	5.43
22	2	6.57	6.85	5.97	6.74
21	3	15.19	15.49	13.93	15.35
20	5	9.58	9.77	8.69	9.67
19	3	6.14	6.26	6.10	6.19
18	2	21.38	21.74	21.22	21.52
17	5	6.87	6.94	6.90	6.90
16	2	.78	1.18	.98	.94
15	3	3.92	3.78	3.67	3.86
13	2	4.14	4.59	3.99	4.37
12	3	8.92	9.39	8.82	9.20
11	1	4.96	4.46	4.51	4.62
9	1	2.94	3.48	1.99	3.46
6	1	7.30	7.89	4.99	8.24

TABLE 8

Average Discrepancy Between CAT14 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (819) - Grade 4, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equi-percentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
52	3	4.70	4.62	8.67	7.43
51	9	3.95	3.92	5.67	5.22
50	8	7.00	6.95	6.00	6.08
49	6	8.15	8.08	7.33	7.33
48	9	6.57	6.59	6.78	6.83
47	14	6.55	6.53	6.33	6.30
46	10	4.48	4.54	4.55	5.00
45	15	5.05	5.06	5.10	5.44
44	15	6.31	6.30	6.30	6.28
43	15	6.30	6.27	6.32	6.26
42	17	5.42	5.44	5.33	5.41
41	11	5.55	5.54	5.73	5.57
40	11	7.22	7.27	6.32	6.94
39	17	7.02	7.04	6.46	6.77
38	15	7.03	7.02	7.66	7.11
37	8	4.34	4.34	4.30	4.33
36	7	6.76	6.76	7.10	6.89
35	17	6.48	6.49	6.44	6.43
34	13	5.98	5.98	6.18	5.91
33	10	6.07	6.06	5.80	5.87
32	6	7.26	7.23	5.53	6.60
31	5	6.91	6.92	7.29	7.11
30	7	5.57	5.58	5.75	5.71
29	6	6.51	6.49	6.13	6.20
28	5	8.82	8.80	8.65	8.64
27	5	7.58	7.52	7.33	7.08
26	10	4.56	4.59	4.63	4.72
25	11	6.43	6.47	6.54	6.63
24	6	9.07	9.12	9.12	9.29
23	3	2.61	2.56	2.77	2.41
22	3	7.61	7.79	6.27	8.15
21	7	4.29	4.32	4.80	4.36
20	6	8.67	8.81	7.33	8.94
19	4	5.50	5.60	5.50	5.65
18	4	9.00	9.00	9.00	9.00
17	8	7.81	7.94	6.38	7.92
16	7	7.28	7.39	6.43	7.34
15	3	7.41	7.32	8.47	7.38
14	4	7.71	8.01	4.75	7.78
13	2	8.66	8.98	8.50	8.70
12	5	8.67	8.86	8.90	8.68
11	1	2.44	2.09	2.50	2.43
10	2	11.01	11.37	10.00	11.02
9	2	8.96	9.34	7.75	8.99
8	1	1.91	2.30	0.00	1.98
2	1	26.61	27.09	16.67	27.42

TABLE 9

Average Discrepancy Between CAT15 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (820) - Grade 5, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
48	5	5.02	5.57	4.10	3.59
47	3	2.71	2.54	3.52	3.12
46	3	12.97	13.47	11.34	11.96
45	6	3.17	3.55	2.56	2.64
44	11	7.01	7.45	5.85	6.41
43	5	2.42	2.21	3.15	2.75
42	10	4.73	4.88	4.37	4.57
41	10	3.82	4.03	3.43	3.64
40	11	7.53	7.73	7.25	7.41
39	13	6.09	6.16	6.05	6.08
38	10	5.07	5.12	5.07	5.07
37	16	3.50	3.50	3.50	3.50
36	8	7.56	7.51	7.45	7.52
35	12	6.69	6.63	6.60	6.61
34	16	6.91	6.94	6.85	6.99
33	11	6.96	7.03	6.53	7.18
32	13	6.43	6.48	6.26	6.65
31	13	5.79	5.76	5.83	5.62
30	8	9.02	9.02	8.89	8.91
29	11	4.59	4.60	3.38	4.13
28	13	7.28	7.29	7.35	7.25
27	10	7.85	7.84	8.13	7.94
26	10	7.88	7.84	8.69	8.06
25	13	6.80	6.81	7.14	6.77
24	17	7.02	7.10	6.31	6.82
23	6	4.17	4.17	4.17	4.17
22	11	7.93	8.06	7.59	7.80
21	7	6.85	7.02	6.02	6.71
20	6	9.55	9.64	9.17	9.51
19	15	6.38	6.40	6.41	6.38
18	11	12.30	12.33	12.28	12.31
17	9	9.01	9.21	8.99	19.14
16	12	5.81	5.88	6.20	5.88
15	7	11.13	11.19	11.60	11.21
14	3	6.84	6.69	5.95	6.59
13	3	8.32	8.16	8.33	8.00
12	7	6.54	6.75	7.31	7.05
11	5	2.23	2.34	2.85	2.52
10	5	5.02	5.36	6.40	6.05
9	2	4.24	3.65	3.50	3.50
8	2	2.82	3.43	5.50	5.16
7	1	2.37	3.02	7.00	5.09

TABLE 10

• Average Discrepancy Between CAT16 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (820) - Grade 6, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY			
		LO	LT	EQ	RM
47	2	4.50	4.50	4.50	6.45
46	8	4.48	4.56	7.63	8.74
45	13	4.82	4.78	5.17	5.61
44	16	5.67	5.78	6.63	7.30
43	16	5.19	5.09	4.88	4.93
42	20	4.25	4.21	4.29	4.63
41	22	5.90	5.94	6.04	6.28
40	25	5.26	5.27	5.25	5.36
39	20	4.77	4.75	4.93	4.75
38	23	5.89	5.92	5.64	5.95
37	18	5.94	5.97	5.66	5.91
36	24	5.46	5.45	5.55	5.49
35	17	5.74	5.74	6.10	5.81
34	15	5.81	5.82	5.36	5.64
33	16	6.51	6.51	6.38	6.38
32	19	5.27	5.27	5.03	5.04
31	11	8.22	8.20	7.84	7.95
30	11	6.68	6.71	7.60	7.29
29	8	4.38	4.38	4.58	4.61
28	8	5.28	5.31	5.43	5.66
27	15	4.78	4.79	4.85	5.06
26	6	5.80	5.86	5.99	6.31
25	9	4.69	4.81	4.86	5.53
24	11	6.78	6.67	6.86	6.10
23	7	4.95	4.92	5.02	4.97
22	8	5.49	5.63	5.29	6.18
21	10	7.24	7.37	6.70	7.77
20	11	5.39	5.42	5.49	5.50
19	2	8.38	8.75	6.80	9.55
18	9	4.28	4.42	3.78	4.64
17	6	10.44	10.73	9.65	11.09
16	5	6.96	7.15	6.66	7.39
15	3	3.85	3.69	4.60	3.84
14	4	8.79	9.05	7.25	9.10
13	2	10.72	11.25	6.99	11.19
11	2	12.00	12.59	10.00	12.18
10	3	9.97	10.58	5.31	10.00
9	2	18.78	19.42	14.50	18.65
8	1	13.92	14.59	10.00	13.62

TABLE 11

Average Discrepancy Between CAT17 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (821) - Grade 7, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
48	4	2.61	2.89	5.00	4.66
47	5	2.88	2.78	3.18	3.04
46	8	4.61	4.97	6.60	6.29
45	7	3.72	3.51	4.53	3.90
44	10	5.72	5.54	5.70	5.21
43	13	6.44	6.35	5.87	6.22
42	15	6.73	6.45	4.66	6.25
41	18	6.54	6.62	7.64	6.63
40	21	6.21	6.32	6.58	6.26
39	15	7.91	7.93	8.00	7.90
38	15	9.53	9.61	9.44	9.48
37	20	8.00	8.07	7.75	7.89
36	15	6.74	6.78	6.39	6.64
35	16	7.60	7.66	7.11	7.38
34	22	7.10	7.06	7.79	7.27
33	16	7.21	7.24	6.68	7.04
32	20	8.13	8.17	7.41	7.91
31	25	5.05	5.06	4.87	4.87
30	20	7.04	7.05	6.55	6.83
29	13	8.34	8.33	7.43	8.03
28	14	3.34	3.35	4.12	3.60
27	15	6.52	6.48	5.98	6.24
26	13	6.86	6.81	6.50	6.59
25	16	6.58	6.65	6.65	6.80
24	19	5.12	5.11	5.11	5.11
23	16	6.38	6.38	6.38	6.38
22	12	6.41	6.38	6.36	6.38
21	16	4.90	4.93	4.87	4.91
20	10	6.36	6.41	6.30	6.36
19	20	5.64	5.69	5.57	5.63
18	12	5.25	5.25	5.25	5.25
17	16	4.70	4.89	4.58	4.59
16	9	6.77	6.99	6.25	6.54
15	8	4.46	4.67	4.00	4.23
14	8	7.14	7.37	6.46	6.99
13	7	4.37	4.85	1.84	3.72
12	10	5.42	5.83	3.49	4.86
11	12	6.27	6.63	5.17	5.76
10	5	6.93	7.50	4.26	6.13
9	6	13.94	14.34	12.22	13.89
7	3	6.08	6.74	2.65	5.31
5	1	8.66	9.37	5.49	8.08

TABLE 12

Average Discrepancy Between CAT18 Converted Scores Yielded by Four Equating Methods and Those Actually Observed for Each Item Bank Raw Score (821) - Grade 8, Mathematics. (LO - Linear Observed; LT - Linear True; EQ - Equipercentile; RM - Rasch Model)

ITEM BANK RAW SCORE	FREQUENCY	AVERAGE DISCREPANCY LO	LT	EQ	RM
46	14	3.95	4.18	5.23	5.37
45	11	3.37	3.28	3.36	3.40
44	22	6.00	6.06	6.31	6.16
43	28	5.45	5.51	5.62	5.53
42	26	4.77	4.77	4.77	4.77
41	22	5.74	5.77	5.70	5.73
40	17	4.89	4.91	4.91	4.91
39	19	4.69	4.70	4.70	4.70
38	16	7.77	7.83	7.68	7.63
37	24	6.07	6.14	5.89	5.85
36	15	4.77	4.74	4.88	4.92
35	27	5.34	5.40	5.25	5.11
34	20	8.57	8.56	8.63	8.65
33	19	6.92	6.92	6.95	6.98
32	23	5.20	5.20	5.27	5.30
31	13	10.87	10.86	10.82	10.59
30	19	7.53	7.54	7.55	7.63
29	15	10.24	10.22	10.13	10.13
28	13	8.94	8.95	8.95	8.98
27	15	7.34	7.32	7.34	7.31
26	13	5.05	5.13	5.00	5.15
25	8	7.13	7.13	7.31	7.13
24	14	6.22	6.29	6.00	6.22
23	19	7.37	7.46	7.24	7.35
22	15	7.19	7.32	6.78	7.10
21	11	9.04	9.01	9.19	9.08
20	11	8.64	9.06	7.54	8.21
19	14	6.72	6.79	6.88	6.71
18	22	8.54	8.70	8.32	8.32
17	8	9.65	9.92	9.46	9.19
16	8	9.70	9.85	9.60	9.45
15	6	12.86	13.06	12.94	12.48
14	10	10.19	10.58	10.61	9.45
13	6	12.22	12.67	14.34	11.33
12	4	8.34	9.05	12.55	6.93
11	1	13.09	13.84	16.47	11.62
10	1	5.85	6.64	7.98	4.33
9	1	14.61	15.43	15.97	13.06
8	2	10.86	11.72	11.98	9.31
4	1	6.39	7.39	6.00	5.07

APPENDIX E

Graphs of the Discrepancy Between the CAT/C Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score

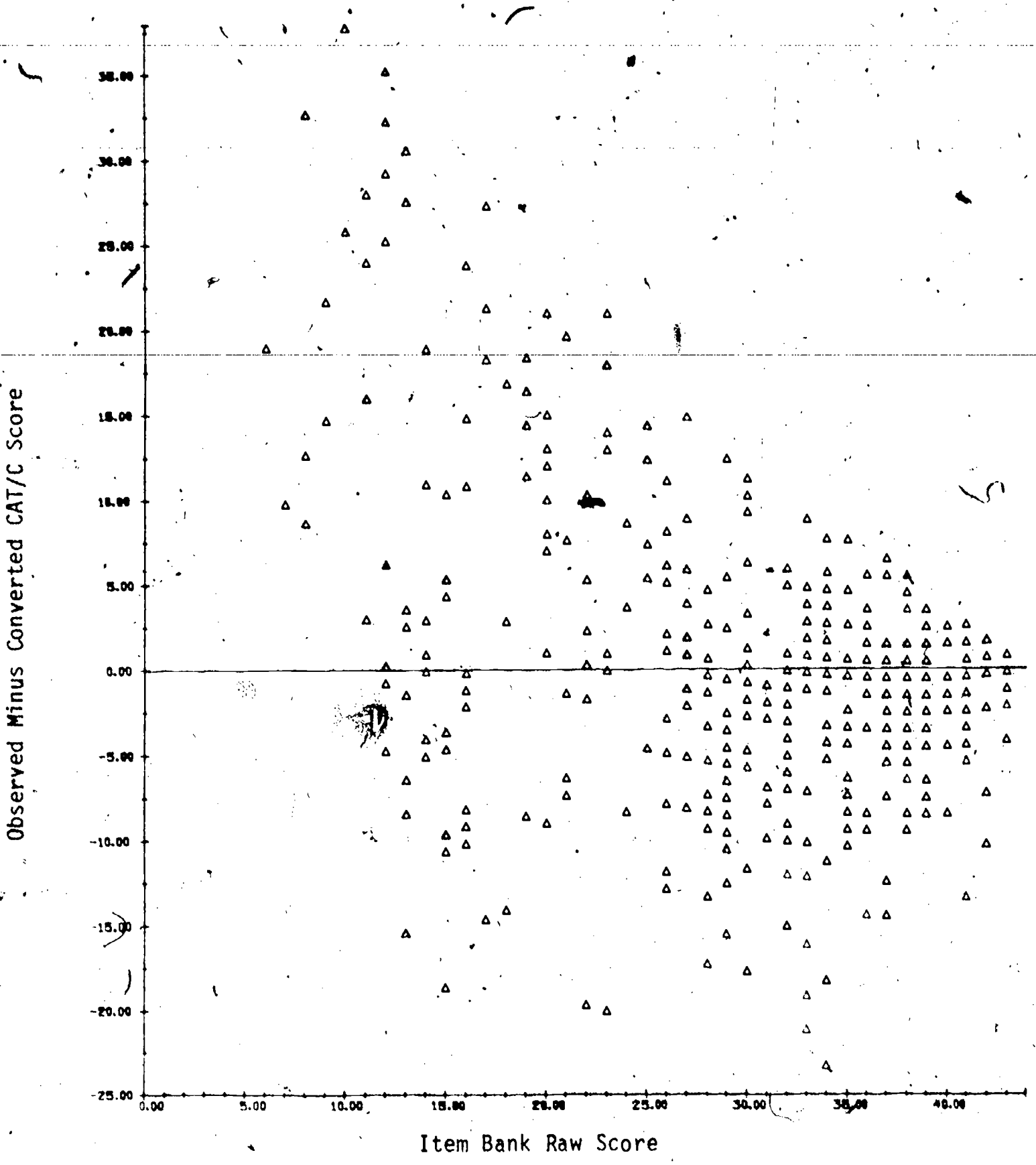

Figure 1

Graph of the Discrepancy Between the CAT13 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (318) - Grade 3, Reading.

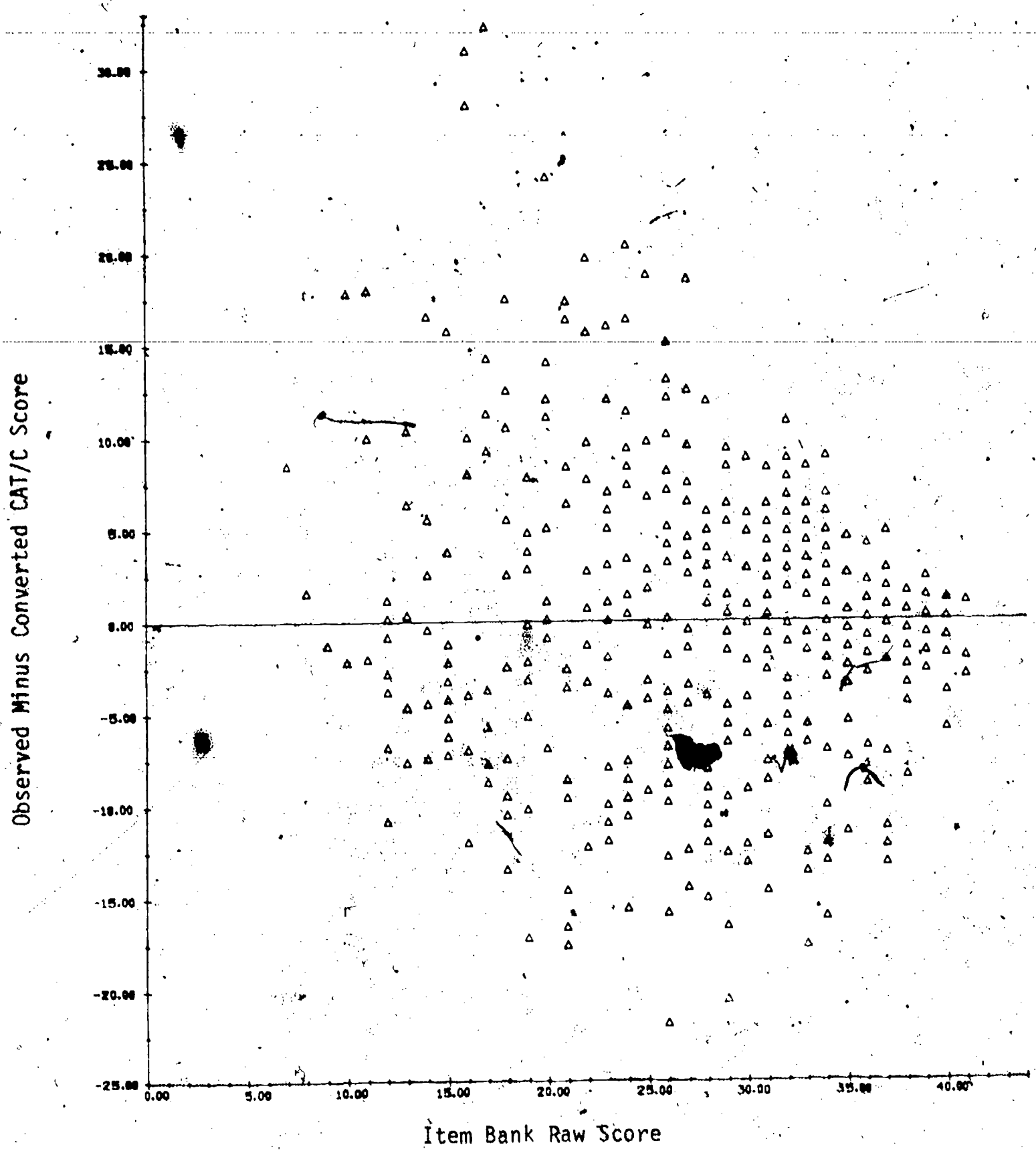

Figure 2

Graph of the Discrepancy Between the CAT14 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (319) - Grade 4, Reading.

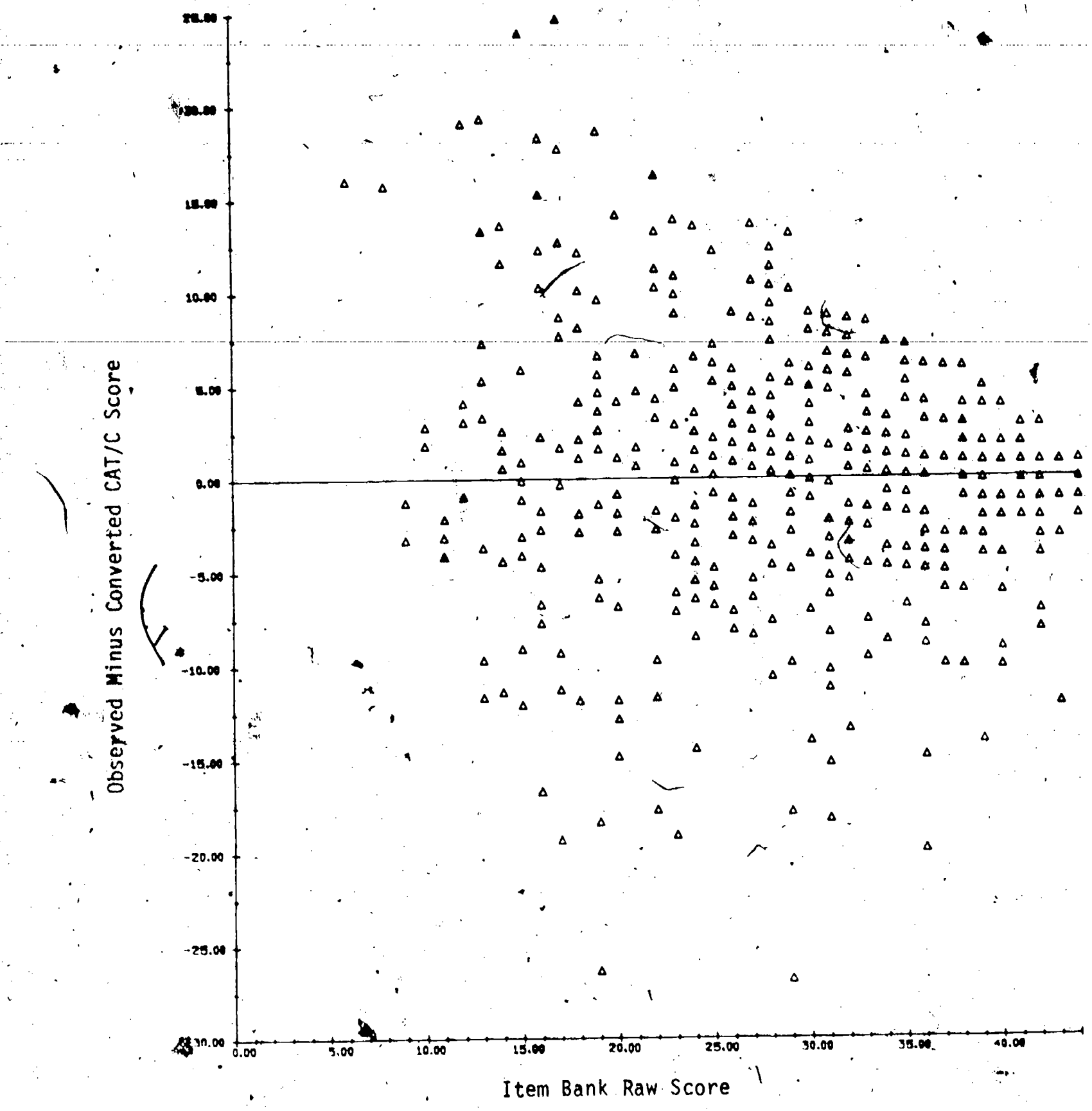

Figure 3

Graph of the Discrepancy Between the CAT15 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (320) - Grade 5, Reading.

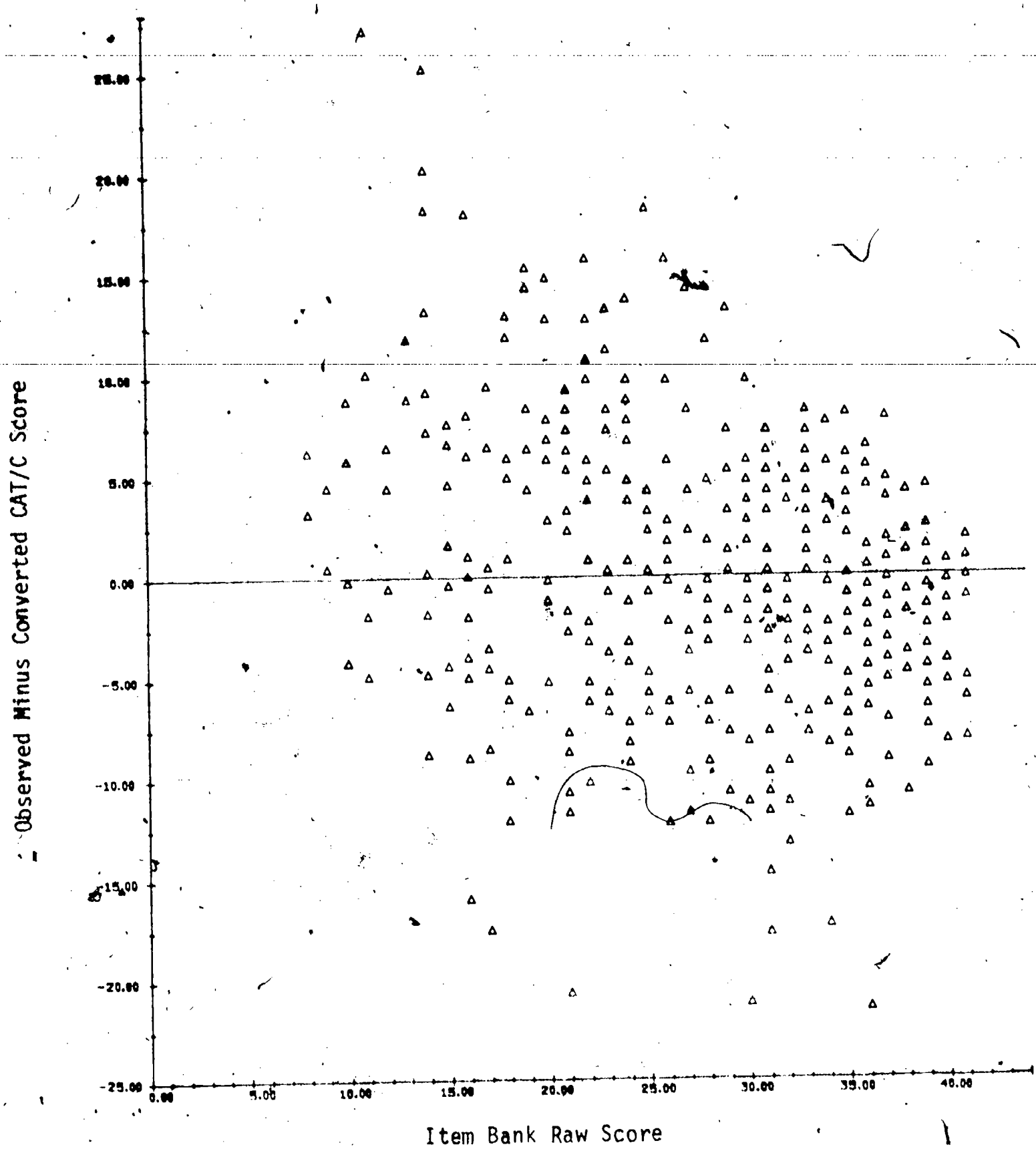

Figure 4

Graph of the Discrepancy Between the CAT16 Raw Score Actually Obtained by Each Student Testeb and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (320) - Grade 6, Reading.

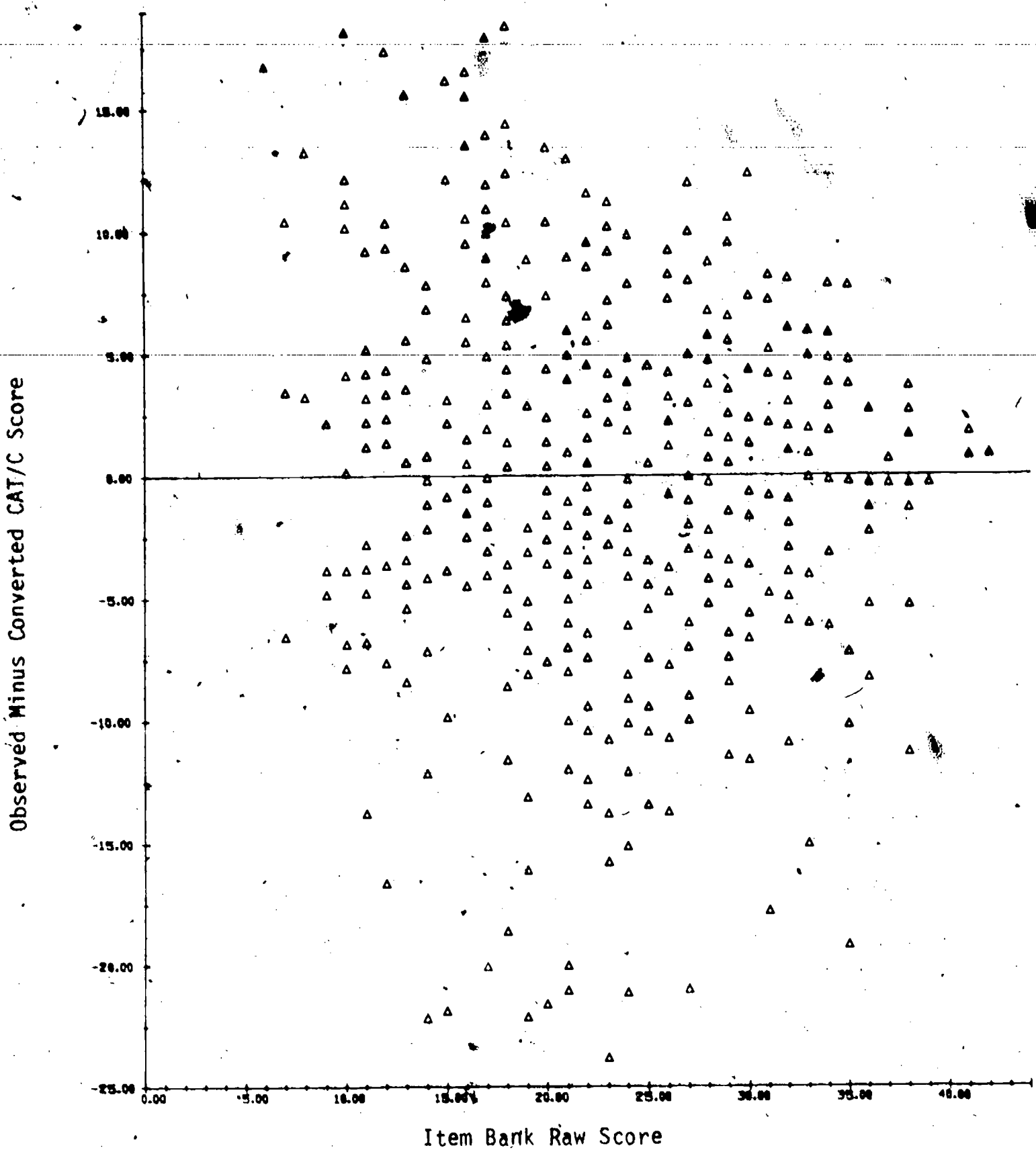

Figure 5

Graph of the Discrepancy Between the CAT17 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (321) - Grade 7, Reading.

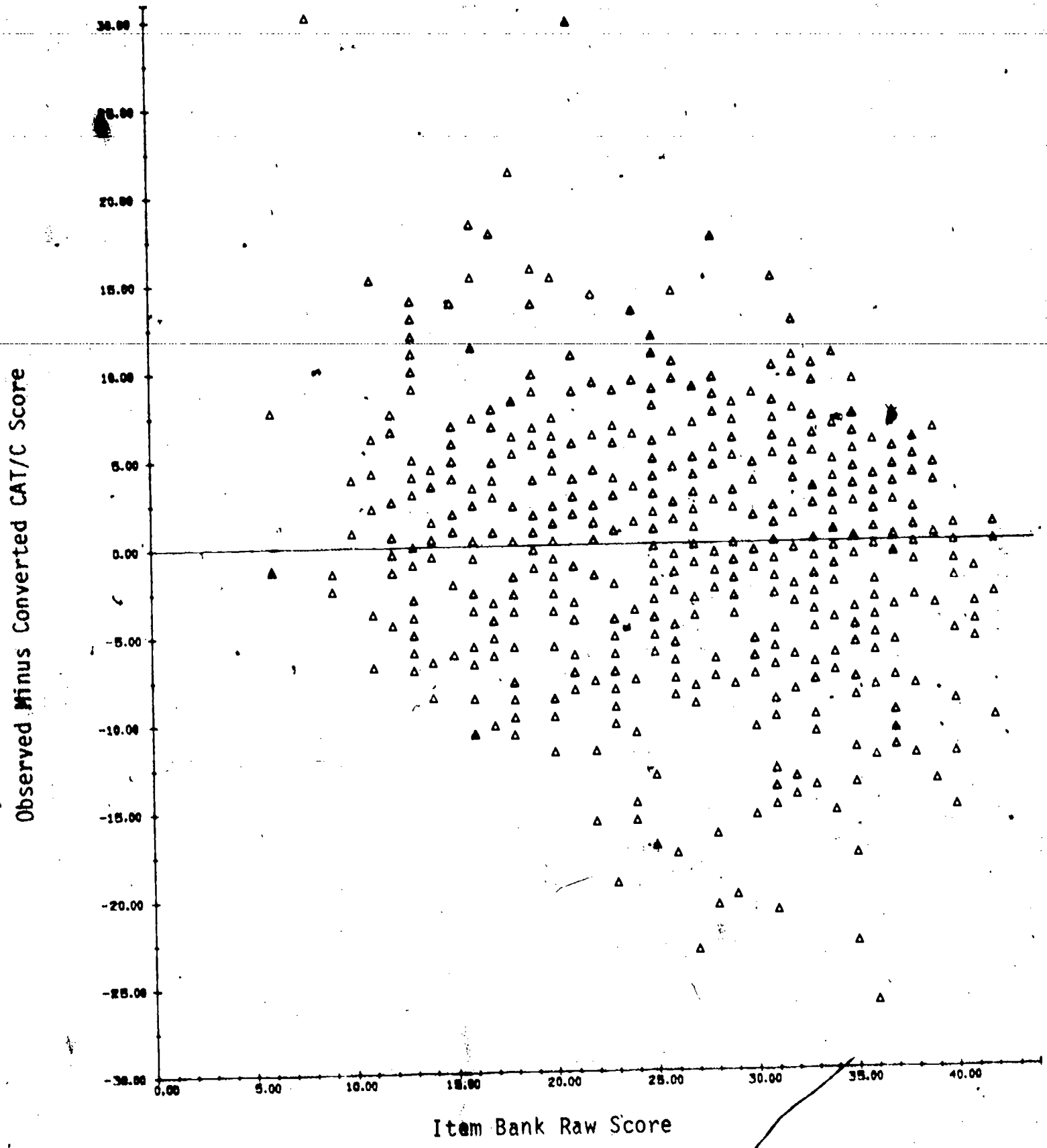

Figure 6

Graph of the Discrepancy Between the CAT18 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (321) - Grade 8, Reading.

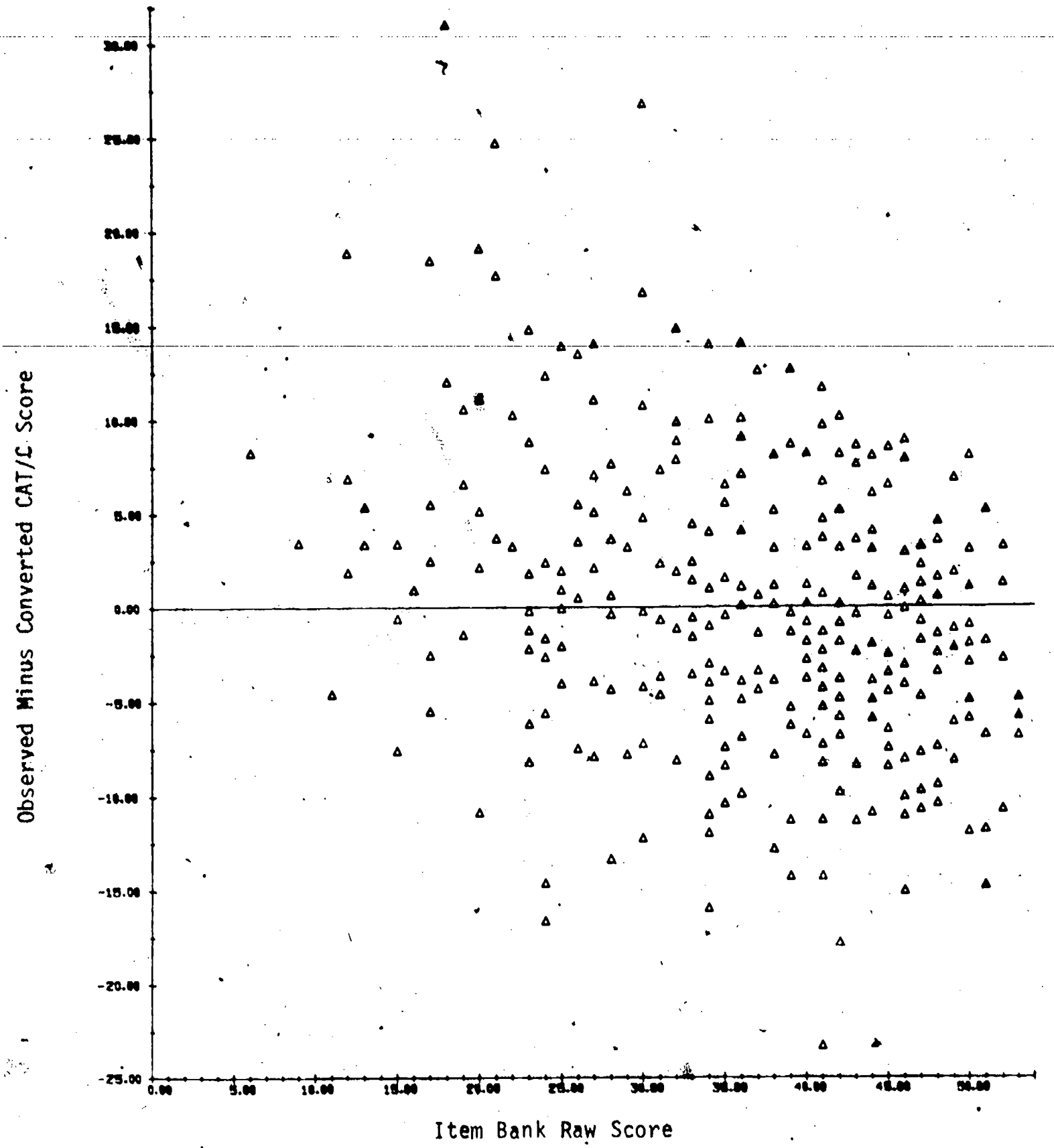

Figure 7

Graph of the Discrepancy Between the CAT13 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (818) - Grade 3, Mathematics.

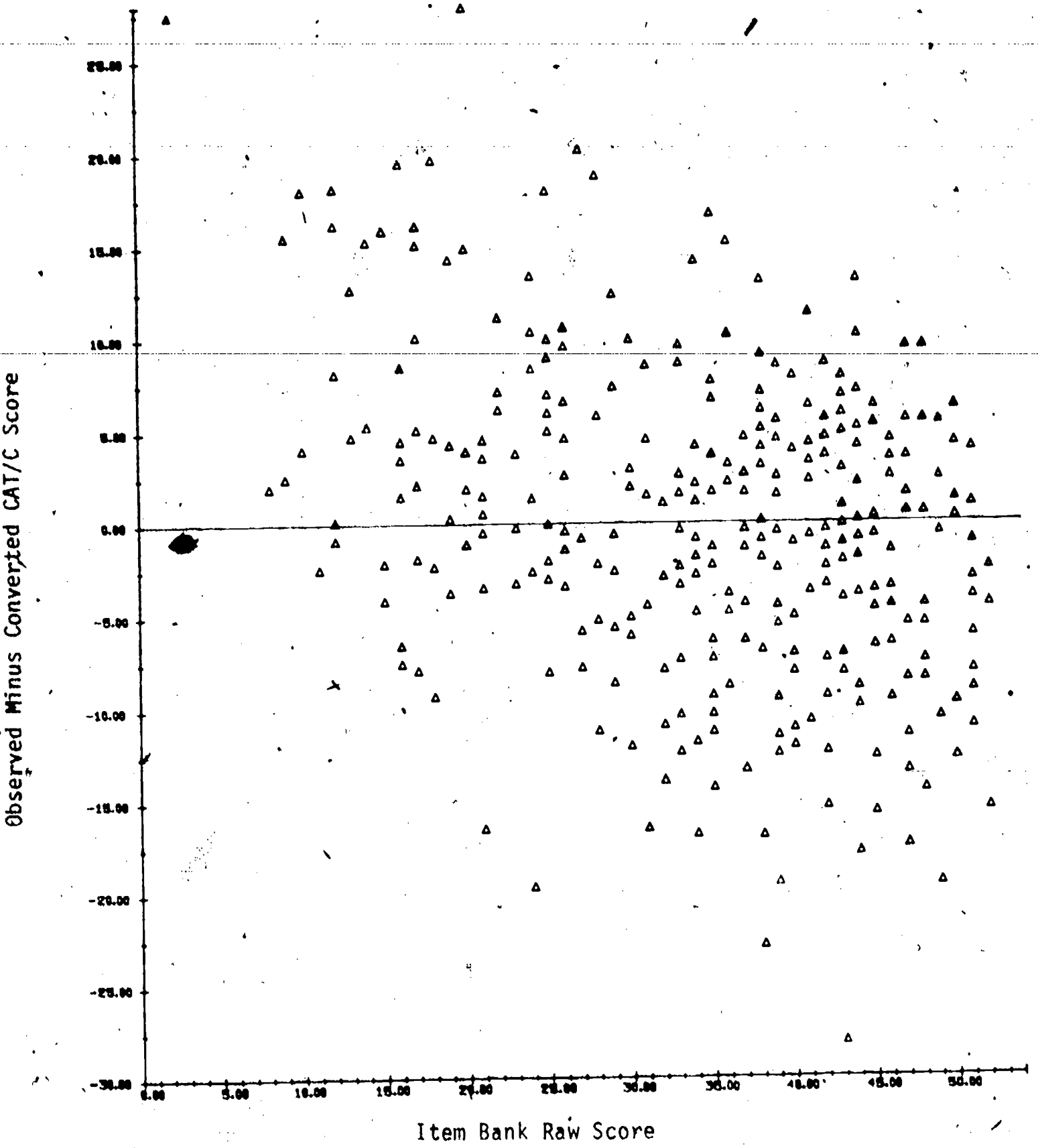

Figure 8

Graph of the Discrepancy Between the CAT14 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (819) - Grade 4, Mathematics.

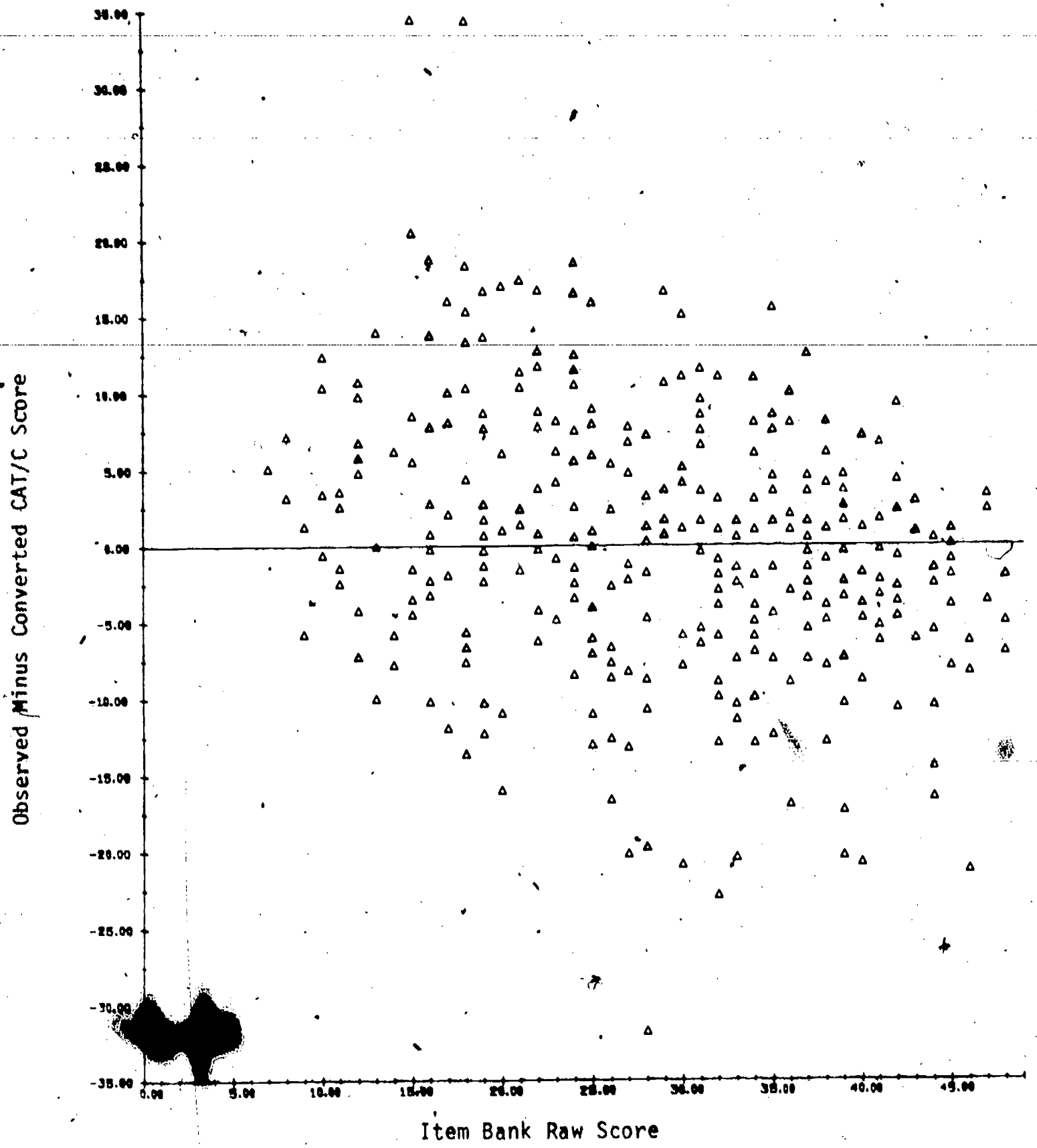

Figure 9

Graph of the Discrepancy Between the CAT15 Raw Score Actually Observed by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (820) - Grade 5, Mathematics.

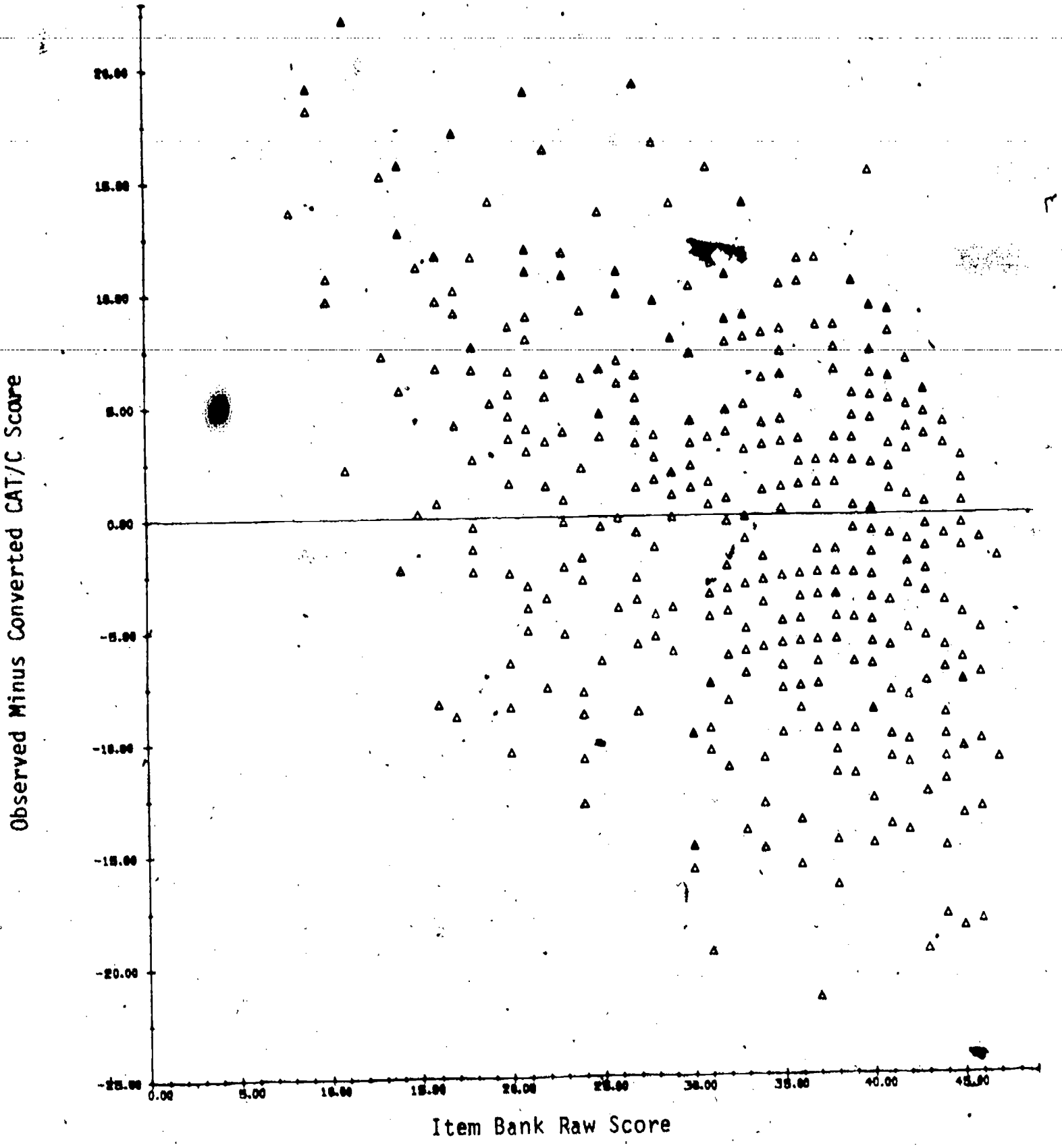

Figure 10

Graph of the Discrepancy Between the CAT16 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (820) - Grade 6, Mathematics.

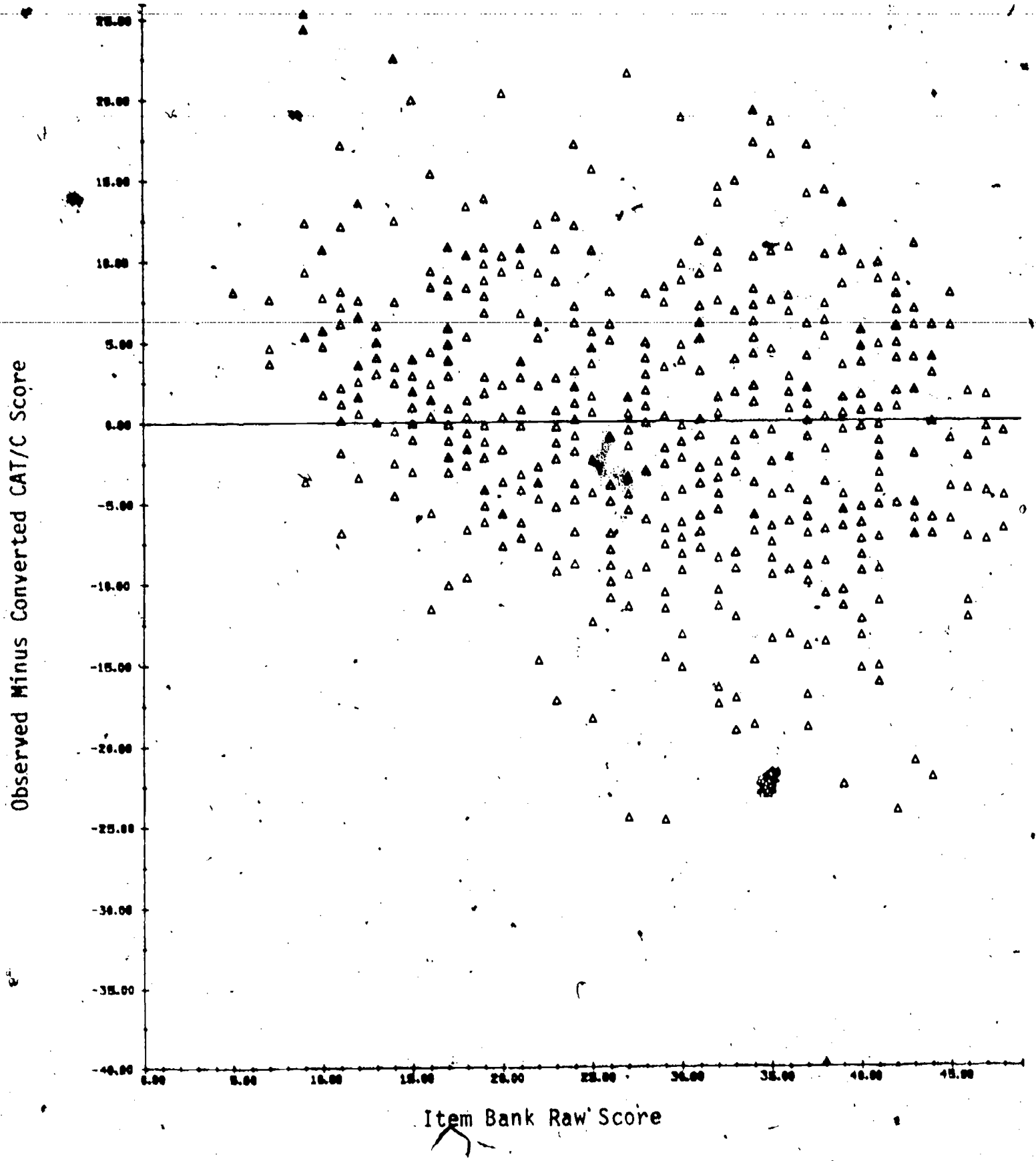

Figure 11

Graph of the Discrepancy Between the CAT17 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (821) - Grade 7, Mathematics.

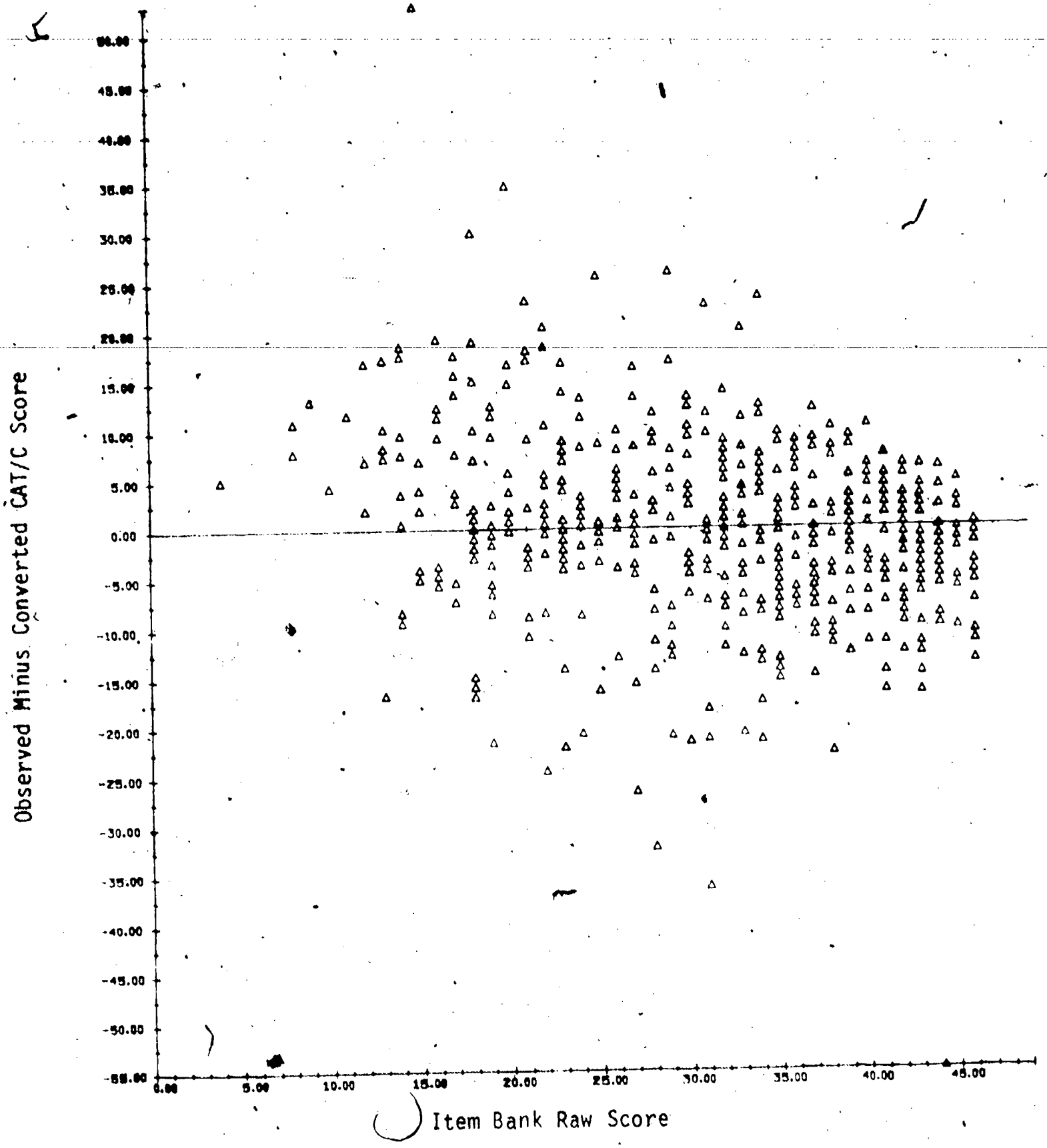

Figure 12

Graph of the Discrepancy Between the CAT18 Raw Score Actually Obtained by Each Student Tested and the CAT/C Converted Score Yielded by the Rasch Model Equating Technique Estimated for Students Obtaining that Item Bank Raw Score (821) - Grade 8, Mathematics.